

Vinícius Barreto Klein

**UMA PROPOSTA DE MODELO CONCEITUAL PARA USO DE
*BIG DATA E OPEN DATA PARA SMART CITIES***

Dissertação submetida ao Programa de Pós- Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina para a obtenção do grau de Mestre em Engenharia e Gestão do Conhecimento.

Orientador: Prof. José L.Todesco, Dr .
Co-orientador: Prof. Fernando Álvaro Ostuni Gauthier, Dr.

Florianópolis
2015

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária
da UFSC.

Klein, Vinícius Barreto
Uma proposta de Modelo Conceitual para uso
de Big Data e Open Data para Smart Cities /
Vinícius Barreto Klein ; orientador, José
Leomar Todesco ; coorientador, Fernando Álvaro
Ostuni Gauthier. - Florianópolis, SC, 2015.
168 p.

Dissertação (mestrado) - Universidade
Federal de Santa Catarina, Centro Tecnológico.
Programa de Pós-Graduação em Engenharia e
Gestão do Conhecimento.

Inclui referências

1. Engenharia e Gestão do Conhecimento. 2.
smart cities. 3. big data. 4. open data. I.
Todesco, José Leomar. II. Gauthier, Fernando
Álvaro Ostuni. III. Universidade Federal de
Santa Catarina. Programa de Pós Graduação em
Engenharia e Gestão do Conhecimento. IV.
Título.

Vinícius Barreto Klein

**UMA PROPOSTA DE MODELO CONCEITUAL PARA USO DE
*BIG DATA E OPEN DATA PARA SMART CITIES***

Esta dissertação foi julgada adequada para obtenção do Título de Mestre em Engenharia e Gestão do Conhecimento, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina.

Florianópolis, 10 de dezembro de 2015.

Prof. Roberto Carlos dos Santos Pacheco, Dr.
Coordenador do Curso

Banca Examinadora:

Prof. José Leomar Todesco, Dr.
Orientador – PPGEGC/UFSC

Prof. Fernando Álvaro Ostuni Gauthier, Dr.
Coorientador – PPGEGC/UFSC

Prof. Rogério Cid Bastos, Dr.
Membro Interno – PPGEGC/UFSC

Prof. Mario Antônio Ribeiro Dantas, Ph.D.
Membro Interno- PPGEGC/UFSC

Prof. Renato Fileto, Dr.
Membro Externo

AGRADECIMENTOS

Agradeço primeiramente a Deus, em sua forma Jesus Cristo, pois sem Ele nada é feito, fonte de minha força e inspiração maior.

Registro um grande agradecimento à minha esposa (Siluana Grassi Klein), pelo apoio incondicional, pois cumpriu com excelência seu papel de auxiliadora e companheira durante o mestrado, com muita compreensão, dedicação e amor. Da mesma forma, ao meu pai (Marcos Vinicius Klein), mãe (Eloisa Barreto Klein), irmão (Lucas Barreto Klein) e demais membros da família, devido ao eterno suporte e incentivo dado a este projeto pessoal e profissional.

Ao orientador, professor e grande inspirador deste trabalho, José Leomar Todesco, que desde a definição do seu tema inicial até sua conclusão, guiou com extremo domínio, sabedoria e amizade. Grande mestre.

Agradeço também aos membros da banca de defesa, pois suas contribuições foram muito valiosas para o presente trabalho.

Aos demais professores, desde a graduação na UFSC até o mestrado, e aos antigos e novos colegas feitos no EGC, um forte agradecimento. Esta caminhada foi um prazer.

“Even the most starless midnight may herald the
dawn of some great fulfillment.”

Martin Luther King (Jr.)

RESUMO

Atualmente vivemos um contexto onde a sociedade produz um alto volume de dados, produzidos pelas mais diversas fontes, em diferentes formatos e esquemas distintos, e de forma cada vez mais veloz. Este fato corresponde ao fenômeno *big data*. Contribuindo com este fenômeno, o movimento dados abertos (*open data*) também adiciona novas fontes de dados produzidos pela sociedade atual. Os dados *big data* e os dados abertos podem servir de insumo para a geração de conhecimento, e as *smart cities* (cidades inteligentes) podem se beneficiar deste processo. Estas cidades representam um conceito que envolve utilizar TICs (tecnologias da informação e comunicação) como meio de melhorar a qualidade de vida nos centros urbanos atuais. Esta ideia motiva-se principalmente pelos diversos problemas enfrentados pelos habitantes destas cidades, como o mal gerenciamento de seus recursos naturais, altos índices de poluição atmosférica, trânsito intenso, taxas de crimes dentre outros, causados principalmente pela alta concentração de pessoas nestes locais. Neste contexto, o objetivo desta dissertação é identificar as principais fontes de dados e suas características, e interligá-las às necessidades das *smart cities*. Foi desenvolvida uma proposta de modelo conceitual para *smart cities*, que utiliza *big data* e *open data* como fonte de dados. Para isso, foi realizada primeiramente uma pesquisa exploratória dos temas relacionados à pesquisa que foram organizados na fundamentação teórica. Em seguida, foram desenvolvidas questões de competência e outras práticas do método OntoKEM para desenvolvimento de ontologias, que guiaram o desenvolvimento do Modelo. Estas questões foram respondidas com auxílio do modelo CESM de Bunge. Em seguida, foi proposto o

Modelo, organizado em camadas, e foi realizada sua verificação em um cenário de uso, onde foram apresentadas discussões, resultados e sugestões futuras.

Palavras-Chave: *big data*, dados abertos, *smart cities*, modelo conceitual

ABSTRACT

We currently live in a context where society produces a high volume of data, generated by various sources, in different formats and different schemes, and in an increasingly fast way. This fact corresponds to the big data phenomenon. Contributing to this phenomenon, the movement open data also adds new data sources produced by today's society. The big data and open data can sources serve as input for the generation of knowledge, and the smart cities can benefit from this process. These cities represent a concept that involves using ICTs (information and communication technologies) as means of improving the quality of life in today's urban center. This idea motivated is mainly by the various problems faced by inhabitants of these cities, such as mismanagement of its natural resources, high levels of air pollution, heavy traffic, crime rates and others problems mainly caused by the high concentration of people in these places. In this context, in order to identify the main data sources and their characteristics, and connect them to the needs of smart cities, a proposal for a conceptual model for smart cities was developed, which uses big data and open data as data sources. To do this, first it was made an exploratory research of issues related to research and then they were organized in the theoretical foundation chapter. Then competency questions have been developed and other ontoKEM practices for developing ontologies, which guided the development of the model. These questions were answered with the aid of Bunge CESM model. Then the model was proposed, organized in layers and checked in a usage scenario, where discussions were presented, so as the results and future researches related.

Keywords: big data, open data, smart cities, conceptual models

LISTA DE FIGURAS

Figura 1: Apresentação do problema.....	18
Figura 2: Modelo de smart cities - características genéricas e seus fatores.....	26
Figura 3: O fenômeno big data.....	32
Figura 4: Países top 20 no Open Data Ranking.....	37
Figura 5: Governança de Dados e Qualidade de Dados segundo DMBOK.....	42
Figura 6: Ciclo mínimo de produção tecnológica.....	57
Figura 7: Aplicação do Modelo CESM às Smart Cities Baseadas em Big Data e Open Data.....	71
Figura 8: Modelo Conceitual.....	72
Figura 9: Camada Conceitual.....	77
Figura 10: Camada de Domínio e Indicadores.....	90
Figura 11: Camada de Dados (Data Layer).....	99
Figura 12: As três camadas do Modelo Conceitual.....	104
Figura 13: Ciclo Virtuoso de Uso e Produção de Dados e Conhecimento nas Smart Cities.....	149

LISTA DE QUADROS

Quadro 1: Definições de smart cities.....	27
Quadro 2: Amostra sobre a composição do Open Data Index.....	35
Quadro 3: Dimensões de qualidade da informação.....	44
Quadro 4: Delineamento da pesquisa.....	58
Quadro 5: Consultas em bases bibliográficas.....	58
Quadro 6: Ilustração através de exemplos de aplicação do modelo CESM de Bunge.....	64
Quadro 7: Aplicação do modelo CESM de Bunge para desenvolver o Modelo.....	69
Quadro 8: Dados estruturados, semi e não estruturados.....	97
Quadro 9: Metadados Dublin Core.....	101
Quadro 10: Formulário para o mapeamento dos dados por domínio: Environment.....	105
Quadro 11: Formulário para o mapeamento dos dados por domínio: Mobility.....	107
Quadro 12: Formulário para o mapeamento dos dados por domínio: Government.....	108
Quadro 13: Formulário para o mapeamento dos dados por domínio: Economy.....	109
Quadro 14: Formulário para o mapeamento dos dados por domínio: People.....	111
Quadro 15: Formulário para o mapeamento dos dados por domínio: Living.....	112
Quadro 16: Dublin Core Aplicado a Smart Cities.....	113
Quadro 17: Formulário para proveniência das fontes de dados big data e open data para smart cities.....	117
Quadro 18: Formulário para o mapeamento dos dados por domínio: Environment.....	121
Quadro 19: Fonte de Dados ELETROPAULO – Consumo de Energia per capita.....	125
Quadro 20: Fonte de Dados SABESP – Uso Eficiente de Água.....	128
Quadro 21: Fonte de Dados Secretaria do Verde e Meio Ambiente - Área Verde.....	131
Quadro 22: Fonte de Dados Secretaria Municipal de Serviços - Reciclagem.....	135
Quadro 23: Fonte de Dados Relatório do Ministério do Meio Ambiente – Opinião e Engajamento Público.....	138
Quadro 24: Fonte de Dados Redes Sociais – Amostra de Postagens..	141
Quadro 25: Informações Obtidas para Análise.....	145

LISTA DE ABREVIATURAS E SIGLAS

API- *Application Programming Interface*;
BSP- *Business Systems Planning*;
CEP- Código de Endereçamento Postal;
CESM - *Component Environment Structure Mechanism*;
CO2 - Gás Carbônico;
CSV - *Comma Separated Value*;
DGA - Dados Governamentais Abertos;
E-Gov - *Eletronic Government*;
ESPON - *European Observation Network for Territorial Development*;
EU - *European Union*;
GD – Governança de Dados
GDP - *Gross Domestic Product*;
GPS - *Global Positioning System*;
HP - *Hewlett-Packard*;
HOS - *Home and Office Automation*;
HQ - *HeadQuarters*;
HTML - *Hiper Text Markup Language*;
IBM - *International Business Machines*;
IBGE - Instituto Brasileiro de Geografia e Estatística;
ICF - *Intelligent Community Forum*;
ICT - *Information and Communication Technologies*;
IEC - *International Electrotechnical Commission*;
IoT - *Internet of Things*;
ISCED - *Intenational Standard Classification of Education*;
ISO - *International Organization for Standardization*;
JSON - *Javascript Object Notation*;
LED - *Light Emitting Diode*;
KW - Kilowatt;
LOD - *Linked Open Data*;
OLTP - *Online Transaction Processing*
OLAP - *Online Analytical Processing*
OGD - *Open Government Data*;
OSN - *Online Social Network*;
OWL - *Web Ontology Language*;
PPS - *Prospective Payment System*;
REST - *Representational State Transfer*;
R&D - *Research and Development*;
RDF - *Resource Description Framework*;

SABESP - Companhia de Saneamento Básico do Estado de São Paulo;

SKOS - *Knowledge Organization Systems*

SGBD - Sistema Gerenciador de Banco de Dados;

SOAP - *Simple Object Access Protocol*;

SVMA - Secretaria do Verde e Meio Ambiente de São Paulo;

TIC - Tecnologia da Informação e Comunicação;

URL - *Uniform Research Locator*;

XML - *Extensible Markup Language*;

SUMÁRIO

1 INTRODUÇÃO.....	17
1.1 APRESENTAÇÃO DO PROBLEMA DE PESQUISA.....	17
1.2 OBJETIVOS.....	19
1.2.1 Geral.....	19
1.2.2 Específicos.....	19
1.3 JUSTIFICATIVA.....	19
1.4 ADERÊNCIA AO EGC.....	23
1.5 CARACTERIZAÇÃO DE PESQUISA.....	23
1.6 ORGANIZAÇÃO DOS CAPÍTULOS.....	24
2 FUNDAMENTAÇÃO TEÓRICA.....	25
2.1 SMART CITIES.....	25
2.2 BIG DATA.....	30
2.3 OPEN DATA.....	34
2.4 GOVERNANÇA DE DADOS.....	40
2.4.1 Qualidade De Dados.....	41
2.4.2 Qualidade Da Informação.....	43
2.5 MODELOS CONCEITUAIS.....	45
2.5.1 Etapas De Construção De Modelos Conceituais.....	47
2.6 TRABALHOS RELACIONADOS.....	48
2.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO.....	53
3 PROCEDIMENTOS METODOLÓGICOS.....	57
3.1 DELINEAMENTO DA PESQUISA.....	57
3.2 PESQUISA BIBLIOGRÁFICA.....	58
4 A PROPOSTA.....	61
4.1 REQUISITOS LEVANTADOS.....	61
4.2 OS PAPÉIS DEFINIDOS E SUAS FUNÇÕES.....	63
4.3 DESENHO DO MODELO E DESCRIÇÃO DE SEU FUNCIONAMENTO.....	71
4.3.1 Camada Conceitual.....	73
4.3.2 Camada De Domínios, Indicadores E Ações.....	78
4.3.3 Camada De Dados.....	91
4.4 CONJUNTO DE ARTEFATOS IMPLEMENTADOS.....	105
4.5 EXPERIMENTO: VERIFICAÇÃO DO MODELO EM UM CENÁRIO DE USO.....	120
5 CONCLUSÕES.....	147
REFERÊNCIAS.....	151

1 INTRODUÇÃO

Neste capítulo são apresentados o problema de pesquisa proposto, os objetivos desta dissertação, suas justificativas, a aderência ao programa de pós-graduação e a caracterização desta pesquisa.

1.1 APRESENTAÇÃO DO PROBLEMA DE PESQUISA

Smart cities é um conceito que envolve diversas características e habilidades que uma cidade deve possuir para melhorar seu desempenho em determinadas áreas, principalmente no gerenciamento de seus serviços e principais recursos (GIFFINGER et al., 2007; VILAJOSANA et al., 2013; NEIROTTI et al., 2014). O adjetivo *smart* caracteriza iniciativas de tornar as cidades atuais mais competitivas, humanas e sustentáveis, aptas a resolver diversos problemas urbanos atuais, como trânsito, segurança pública, poluição dentre outros (GIFFINGER et al., 2007, NAM; PARDO, 2011). Uma destas habilidades envolve utilizar TICs (tecnologias da informação e comunicação) para melhorar a qualidade de vida da população (CARAGLIU; BO; NIJKAMP, 2011).

Segundo Fan e Bifet (2012), *big data* é um fenômeno que vem sendo estudado por diversos pesquisadores tanto da academia como da indústria, e, segundo eles, citando Parker (2012) e Gopalkrishnan (2012), representa uma área com diversos desafios futuros, devido à natureza de seus dados: volumosos e evolutivos. Beyer e Laney (2012) afirmam que *big data* exige novas formas de tratamento tecnológico. Isso se deve principalmente a seu alto volume e velocidade de produção, mas também por possuir diversos formatos e esquemas distintos (DAVENPORT, 2014; O'REILLY, 2012; BEYER; LANEY, 2012; FAN; BIFET, 2012). Em relação ainda ao *big data*, segundo Fan e Bifet (2012), este fenômeno representará uma fronteira a ser vencida por pesquisas científicas e aplicações de negócio. Sua mineração possui grande potencial para a descoberta de novos conhecimentos e geração de *insights*, o que pode capitalizar grande valor em formas de descobertas científicas e diferenciais estratégicos para empresas e outras organizações. Esta criação de novos conhecimentos, em forma de padrões a serem identificados, ou indicadores de desempenho por exemplo, pode auxiliar em muito também as *smart cities* em seu objetivo de melhorar seus processos. Conforme as pesquisas feitas para esta dissertação, estas cidades podem utilizar-se de *big data* para atingir seus objetivos. Dobre e Xhafa (2013) afirmam que são nas cidades que

big data tem seu maior impacto e as *smart cities* devem se basear neste fenômeno para melhorar a qualidade de vida de seus cidadãos. No entanto, implantar uma *smart city* a partir de *big data* é um processo que envolve diversos desafios a serem considerados, tanto financeiros, administrativos como tecnológicos (VILAJOSANA et al., 2013). Do ponto de vista tecnológico, é necessário considerar os diversos domínios de atuação de uma *smart city* (GIFFINGER et al., 2007; NEIROTTI et al. 2014), analisar quais dados são necessários, quais fontes utilizar e ainda como medir e analisar os resultados, de forma que tragam conhecimentos de valor para estas cidades melhorarem seu desempenho. Esta problemática é o foco desta dissertação, conforme Figura 1. Segue abaixo uma lista que objetiva ilustrar as perguntas de pesquisa realizadas nesta dissertação, para melhor entendimento:

Figura 1: Apresentação do problema



Fonte: o autor.

Considerando o contexto das *smart cities*, cujo principal objetivo é melhorar a qualidade de vida nos centros urbanos atuais, e considerando *big data* e *open data* como fonte de dados para estas cidades, seguem as perguntas de pesquisa:

- Como gerar conhecimento para apoiar as *smart cities*, utilizando *big data* e *open data*?
- Quais dados devem ser coletados e quais informações devem

ser produzidas para cumprir este objetivo?

- Quais fontes de dados utilizar? Quais são os principais atores neste processo?

1.2 OBJETIVOS

1.2.1 Geral

Propor um modelo conceitual para *smart cities* utilizando *big data* e *open data* como fontes de dados.

1.2.2 Específicos

- Apresentar o estado da arte nos temas *Big Data*, *Open Data* e *Smart Cities*;
- Especificar o escopo de atuação das *Smart Cities*;
- Elaborar Questões de Competência para guiar o desenvolvimento do Modelo;
- Especificar um conjunto de metadados e mapeamento de fontes de dados para *Smart Cities*;
- Propor o Modelo Conceitual;
- Verificação do Modelo por meio da aplicação em um cenário de uso.

1.3 JUSTIFICATIVA

As cidades atuais em geral, e principalmente as que formam grandes centros urbanos ou metrópoles, possuem diversos problemas que afetam a qualidade vida de seus habitantes. Dobre e Xhafa (2013) apontam que a grande concentração de pessoas nestes locais, e consequentemente, os altos esforços para facilitar as suas condições de vida (exemplos: poluição, segurança e mobilidade urbana) são as principais razões para que uma cidade busque se tornar uma *smart city*. Além disso, apesar de ser um problema típico de grandes metrópoles, cidades de todos os portes podem vir a sofrer com estes problemas,

dependendo de como for realizada sua administração e o quanto íngreme for sua taxa de crescimento. Logo, a abordagem das cidades inteligentes (*smart cities*) tem grande importância e pode auxiliar a vencer tais questões. Este conceito representa a ideia de que estes centros urbanos devem procurar resolver seus problemas através de um melhor gerenciamento de seus recursos ambientais, da mobilidade urbana, de maiores investimentos em capital humano, transparência governamental dentre outros fatores. Para isto, recomenda-se a utilização de TICs (tecnologias da informação e comunicação) como forma de qualificar seus serviços prestados nestas áreas. O estudo deste conceito aponta que determinadas informações sobre os domínios específicos de atuação das *smart cities* (GIFFINGER et al., 2007), se coletadas e analisadas, podem servir de suporte que estas cidades atinjam seus objetivos. Este é o foco desta dissertação.

Somado a esta questão, ao utilizar TICs para resolver os problemas típicos dos centros urbanos atuais, as *smart cities* consequentemente produzem (e consomem) *big data* (DOBRE; XHAFA, 2013). Os sensores e a internet das coisas, somados a outras fontes de dados, como redes sociais e *open data* por exemplo, são grandes contribuidores do aumento de velocidade de produção, volume e diversos formatos de dados disponíveis para estas cidades, constituindo o fenômeno *big data* (GROBELNIK, 2012). De forma geral, estes dados podem ser minerados e fornecer *insights* e conhecimentos valiosos para diversas organizações, incluindo as próprias cidades (FAN; BIFET, 2012). Utilizar *big data* como fonte de conhecimento para as *smart cities* vai de encontro ao conceito de que estas cidades devem utilizar de TICs para aumentar a qualidade de vida de seus habitantes e prover desenvolvimento econômico sustentável, além de conseguir capturar dados, transformar em informação e adaptar suas ações em tempo real (CARAGLIU, BO; NIJKAMP, 2011; TOWNSEND, 2013). Corroborando esta ideia, Dobre e Xhafa (2013) afirmam que são nas *smart cities* que *big data* tem seu real impacto e que os administradores destas cidades precisam de ferramentas para conseguir captar estes dados e reformular suas ações. No entanto, este não é um processo trivial, e possui grande desafio tecnológico, devido a sua complexidade (VILAJOSANA et al., 2013).

Dentro do fenômeno *big data*, optou-se pelas fontes *open data*, escolha esta que se justifica por diversos fatores. O movimento *Open Data* (dados abertos), especialmente no que diz respeito a dados governamentais, traz diversos benefícios para a sociedade, como maior transparência e controle social, mais participação popular, geração de

novos serviços, geração de conhecimento e inovação, além de trazer maior atenção dos diversos setores da sociedade para os problemas municipais (exemplo: epidemias, crimes) (FOUNDATION, [20--]; SCHRIER, 2014). Estas vantagens podem ser associadas por exemplo aos domínios *Smart Governance* (transparência), *Smart People* (participação na vida pública) e *Smart Economy* (empreendedorismo), respectivamente, que, conforme Giffinger et al. (2007), são eixos de atuação das *smart cities*.

Nesta lógica, pode-se pensar que o fato de utilizar *open data* incentiva aumentar estas qualidades (*Smart Governance*, *Smart Economy* e *Smart Economy*) nestas cidades. Além disso, pode inclusive trazer retorno financeiro para o seu governo e sociedade, na medida em que novos empreendimentos e oportunidades de negócio podem ser criados, com o retorno vindo na forma de impostos para o governo e novas oportunidades de negócio para a sociedade. Governos como o dos EUA (U.S. GOVERNMENT, 2015), do Reino Unido (REINO UNIDO, 2015), Brasil (BRASIL, 2015) e diversos outros tem investido nesta iniciativa e alguns retornos interessantes já podem ser notados. Serviços criados pela sociedade (desenvolvedores de software) que utilizam os dados públicos e abertos são um exemplo disso, como o *FlyOnTime.us* por exemplo, que combina dados públicos sobre condições climáticas e vôos (MAYER-SCHONBERGER; CUKIER, 2013), ou os exemplos disponíveis no portal de dados abertos no Brasil¹, e em outros portais, como o lodKEM (LODKEM, 2015) do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento da Universidade Federal de Santa Catarina (PPEGC/UFSC). Conforme Shadbolt et al. (2012), este potencial ainda pode ser aumentado se forem utilizadas técnicas e padrões da web semântica, como *linked data* (dados ligados) e *RDF* (*Resource Description Framework*) por exemplo, para interligar estes dados com outras bases disponíveis, em formatos abertos (não proprietários) e utilizar ligações (*links*) semânticas (passíveis de interpretação por computadores e não apenas por humanos) (BERNERS-LEE, 2006), no que os autores chamam uma combinação de *Linked Data Web* e *Open Government*, como pode ser visto no trabalho *Linked Open Government Data* (SHADBOLT et al., 2012).

Outro aspecto a destacar é que o uso de dados abertos vai de encontro à recente Lei de Acesso à Informação (Lei nº 12.527, de 18 de novembro de 2011), cuja uma das preocupações (ou benefícios) pode-se perceber ser o compartilhamento de conhecimento na esfera pública

1 <http://dados.gov.br/aplicativos/>

brasileira (CARVALHO, 2014). Tal abordagem está alinhada à forma de governo chamada *Open Government*, que sustenta que os cidadãos têm o direito de acesso aos documentos e procedimentos do governo para permitir uma supervisão pública eficaz (LATHROP, RUMA, 2010). Schrier (2014) ilustra seus principais benefícios:

Cada vez mais os cidadãos vão se engajar politicamente [...]; Grupos de defesa do cidadão e pesquisadores serão capazes de analisar dados governamentais e produzir novos e melhores *insights* sobre problemas difíceis (das cidades) de se tratar [...]; Novos negócios (empreendimentos) poderão usar os dados abertos governamentais, combinar com informações de diversas fontes, e produzir novos serviços e produtos, tais como aplicativos para dispositivos móveis; Serviços públicos existentes podem ser melhorados significativamente a partir de análises sobre os dados operacionais, tornando-se possível melhorar seus processos e encurtar seus prazos de execução (SCHRIER, 2014, tradução nossa).

Complementando as justificativas supracitadas, os grandes centros urbanos atuais necessitam se tornar mais otimizados na forma como prestam os seus serviços devido também a sua, em média, baixa performance no gerenciamento de seus recursos naturais. No mundo inteiro, enquanto apenas 50% da população vive em áreas urbanas, (e na Europa por exemplo esta taxa deve chegar aos 80% em 2020), estes centros urbanos são responsáveis por até 80% do consumo de energia mundial, emitindo grandes taxas de CO₂ na atmosfera, dentre outros problemas urbanos causados pela alta concentração de pessoas (ALBINO; BERARDI; DANGELICO, 2015). E no Brasil não é diferente. De acordo com o IBGE (2014), entre os anos de 2013 e 2014, os municípios brasileiros com maior taxa de crescimento foram os que configuram importantes centros regionais ou integravam as principais regiões metropolitanas do país. Além da poluição, os custos com transporte, saúde e segurança também são logicamente maiores nestes centros. Usando a grande Florianópolis como exemplo, segundo relatório de orçamento público (FLORIANÓPOLIS, 2012), entre janeiro e outubro de 2012 foram gastos R\$ 9.729.000,00 com transporte (coletivos urbanos, rodoviários), R\$ 13.714.974,00 com saúde e R\$ 53.126.661,49 com segurança pública. Estes valores podem ser otimizados na medida em que se gerencia com melhor performance

estas áreas, utilizando TICs, informações baseadas no funcionamento das cidades e seus serviços, e investindo em capital humano e social por exemplo, como indica o conceito de *smart cities*.

1.4 ADERÊNCIA AO EGC

Segundo Davenport e Prusak (1998), o conhecimento é uma mistura fluida de experiência condensada, valores, informação contextual e *insight* experimentado, a qual proporciona uma estrutura para avaliação e incorporação de novas experiências e informações. O conhecimento é o principal objeto de estudo do PPGEHC, bem como desta dissertação, cujo foco está em como gerar conhecimento no escopo das *smart cities*, a partir de determinadas fontes de dados (*big data* e *open data*). Pretende-se nesta dissertação tratar o problema proposto através da avaliação (análise) das novas informações geradas através deste modelo, que possibilitará inclusive, novos *insights* sobre o tema, de grande valor para administradores de cidades, bem como demais analistas e/ou pesquisadores deste tema. Além destas justificativas, o conhecimento necessário para melhorar a qualidade de vida nestas cidades (*smart cities*) passa por diversas áreas de conhecimento, o que aumenta a característica interdisciplinar desta pesquisa.

Complementando a aderência, conforme EGC (2015), a linha de pesquisa “Engenharia do Conhecimento Aplicada a Governo Eletrônico” estuda metodologias e técnicas da Engenharia do Conhecimento para a produção de plataformas de governo eletrônico. Logo, um modelo conceitual para uso de dados abertos governamentais para administradores de cidades se encaixa neste contexto. Dentre alguns trabalhos similares no PPEGC, destacam-se Pizzol (2014), que propõe utilizar a *Web of Data* (web de dados) como fonte de informação para o processo de inteligência competitiva setorial, Costa (2013), que propõe um sistema de conhecimento para apoiar a gestão de recursos de programas governamentais. Também menciona-se o trabalho de Rautenberg (2009), que, para mapear as ferramentas da Gestão do Conhecimento e de Agentes Computacionais, utiliza técnicas da Engenharia do Conhecimento, (mais precisamente do desenvolvimento de Ontologias) também utilizadas nesta dissertação.

1.5 CARACTERIZAÇÃO DE PESQUISA

Conforme Gil (2002), uma pesquisa é um procedimento racional e sistemático que tem como objetivo proporcionar respostas aos

problemas que são propostos. Ainda segundo o autor, é possível classificar uma pesquisa realizada mediante a diversos critérios. Essa tarefa auxilia na construção dos procedimentos metodológicos a serem implementados.

Sendo assim, a primeira classificação deste trabalho se faz conforme sua essência, contexto e principalmente finalidade. Ele caracteriza-se como uma pesquisa tecnológica, pois está preocupada em produzir um artefato funcional como resultado final (CUPANI, 2006; FREITAS JUNIOR et al., 2012), baseando-se no conceito de tecnologia de Bunge (1985), a saber:

[...]O estudo científico do artificial [...] Tecnologia pode ser vista como o campo do conhecimento relativo ao projeto de artefatos e ao planejamento de sua realização, operação, ajuste, manutenção e monitoramento, a luz do conhecimento científico (BUNGE, 1985).

Quanto aos procedimentos técnicos adotados, esta pesquisa classifica-se como exploratória/bibliográfica (GIL, 2002). Ela utiliza-se de diversas fontes bibliográficas como consulta para gerar uma melhor familiaridade com o problema, e as principais fontes são periódicos, livros e relatórios técnicos. O delineamento utilizado para esta pesquisa pode ser visto no Capítulo 3.

1.6 ORGANIZAÇÃO DOS CAPÍTULOS

O Capítulo 2 traz a fundamentação teórica dos temas envolvidos (*Smart Cities*, *Big Data*, *Open Data* e Modelos Conceituais), assim como trabalhos relacionados. O Capítulo 3 apresenta os procedimentos metodológicos da pesquisa. No Capítulo 4 é apresentado o Modelo desenvolvido, descrevendo seus requisitos, conjunto de artefatos produzidos e sua verificação através de um experimento. O Capítulo 5 descreve as conclusões e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Na sequência são levantados os fundamentos teóricos necessários para o desenvolvimento desta proposta, conforme segue.

2.1 SMART CITIES

O termo *smart cities* tem estado em evidência nos últimos anos. Segundo Townsend (2013, tradução nossa), *smart cities* são cidades que conseguem capturar dados, transformar em informação e adaptar suas ações em tempo real. Caragliu, Bo e Nijkamp (2011) descrevem uma *smart city* como uma cidade onde se investe em capital humano e social, utilizando-se da infraestrutura de tecnologia de informação para promover o crescimento econômico sustentável e a qualidade de vida. No entanto, nem todas as definições envolvem apenas o uso de TICs. Além destes aspectos, Zygiaris (2012) define uma *smart city* como sendo uma cidade que possui, por exemplo, uma infraestrutura urbana voltada à proteção ambiental e capacidade de se basear no conhecimento e no capital humano e criativo para gerar inovação, aumentando assim sua competitividade econômica. Um trabalho que contribui bastante para o entendimento deste tema foi realizado pela Vienna University, que faz um estudo geral do estado da arte das definições existentes sobre *smart cities*. Os autores relatam que não há uma definição que considere apenas um único aspecto, mas que todas as definições encontradas na literatura envolvem um conjunto de características e habilidades que as cidades devem possuir. Isso inclui desde ser um distrito de TI (tecnologia da informação) até possuir elevado nível de educação de seus habitantes (GIFFINGER et al., 2007). Com base nesta pesquisa eles então abstraem uma série de qualidades encontradas e criam um modelo de características genéricas encontradas nas definições, criando para cada característica um grupo de fatores com indicadores de performance. A Figura 2 apresenta este modelo.

Seguindo o modelo de Giffinger, com base principalmente em dados do Eurostat (escritório de estatística da união européia), em trabalhos de pesquisa do ministério de cultura e comunicação de Paris e do programa ESPON - *European Observation Network for Territorial Development*, os autores então constroem um *ranking* que mostra a performance de um grupo de cidades de porte médio da Europa (eles escolhem este perfil de cidade justificando que as principais pesquisas

urbanas focam em grandes metrópoles, mesmo com a maioria da população urbana vivendo em cidades de médio porte). Este *ranking* nos mostra como podem ser usados estes indicadores para medir a performance destas cidades, provendo uma maneira de avaliar quais áreas é preciso melhorar em uma determinada cidade.

Figura 2: Modelo de *smart cities* - características genéricas e seus fatores.

SMART ECONOMY (Competitiveness) <ul style="list-style-type: none"> ▪ Innovative spirit ▪ Entrepreneurship ▪ Economic image & trademarks ▪ Productivity ▪ Flexibility of labour market ▪ International embeddedness ▪ <i>Ability to transform</i> 	SMART PEOPLE (Social and Human Capital) <ul style="list-style-type: none"> ▪ Level of qualification ▪ Affinity to life long learning ▪ Social and ethnic plurality ▪ Flexibility ▪ Creativity ▪ Cosmopolitanism/Open-mindedness ▪ Participation in public life
SMART GOVERNANCE (Participation) <ul style="list-style-type: none"> ▪ Participation in decision-making ▪ Public and social services ▪ Transparent governance ▪ <i>Political strategies & perspectives</i> 	SMART MOBILITY (Transport and ICT) <ul style="list-style-type: none"> ▪ Local accessibility ▪ (Inter-)national accessibility ▪ Availability of ICT-infrastructure ▪ Sustainable, innovative and safe transport systems
SMART ENVIRONMENT (Natural resources) <ul style="list-style-type: none"> ▪ Attractivity of natural conditions ▪ Pollution ▪ Environmental protection ▪ Sustainable resource management 	SMART LIVING (Quality of life) <ul style="list-style-type: none"> ▪ Cultural facilities ▪ Health conditions ▪ Individual safety ▪ Housing quality ▪ Education facilities ▪ Touristic attractivity ▪ Social cohesion

Fonte: Giffinger et al. 2007.

Ainda nesta linha, Neirotti et al. (2014) fazem um interessante trabalho sobre as tendências atuais em *smart cities*. Os autores se baseiam em Giffinger et al. (2007) e corroboram o conceito afirmando

que as *smart cities* não abrangem apenas inovações tecnológicas, mas também investimentos em capital humano e mudanças no modo de vida destas cidades. Nesta linha, eles classificam os domínios que uma *smart city* deve se preocupar em dois tipos: *hard e soft*. O primeiro diz respeito às áreas em que as TICs tem maior impacto, como em redes de energia, iluminação pública, fontes de energia renováveis, recursos hídricos, lixo, construções inteligentes (todas estas pensando na questão ambiental, evitando desperdícios de recursos através do uso de sensores e sistemas de informação adaptáveis ao contexto). Mobilidade urbana, saúde (utilizando TICs para oferecer melhor atendimento às pessoas) e segurança pública também entram nesta categoria. Já o segundo grupo (*soft*) é onde os sensores e informações adaptadas em tempo real tem menor impacto, pois são áreas que necessitam mais de ações (incentivos) que criem as condições necessárias para que elas se desenvolvam. São encaixadas neste grupo a economia (incentivo ao empreendedorismo, inovação, integração com mercado internacional), administração pública (digitalização da administração pública, *e-gov*, transparência digital para atrair o envolvimento dos cidadãos), educação (uso de TICs para criar mais acesso à educação, políticas educacionais), cultura (mais eventos, monumentos turísticos), inclusão social (diminuição de barreiras para melhorar a qualidade de vida para idosos e pessoas com deficiência, por exemplo).

Mesmo em pesquisas mais recentes, conforme Albino, Berardi e Dangelico (2015) , citando O'Grady e O'Hare (2012), o termo *smart city* ainda não possui uma definição única e completa, que englobe todos os aspectos encontrados nas suas diversas conceitualizações existentes na literatura. Para ilustrar este fato, os autores elencam uma série de definições sobre o termo, conforme Quadro 1. Os autores justificam a dificuldade de formalizar uma única definição pelo fato deste conceito envolver e ser estudado por diversas áreas distintas, cada uma evidenciando aspectos importantes e diversificados.

Quadro 1: Definições de *smart cities*

Conceito
<p>“<i>Smart cities are the result of knowledge-intensive and creative strategies aiming at enhancing the socio-economic, ecological, logistic and competitive performance of cities. Such smart cities are based on a promising mix of human capital (e.g. skilled labor force), infrastructural capital (e.g. high-techcommunication facilities), social capital (e.g. intense and open network linkages) and entrepreneurial capital (e.g. creative and</i></p>

Conceito
<i>risk-taking business activities”. (KOURTIT; NIJKAMP, 2012)</i>
<i>“Smart cities have high productivity as they have a relatively high share of highly educated people, knowledge-intensive jobs, output-oriented planning systems, creative activities and sustainability-oriented initiatives.” (KOURTIT et al., 2012)</i>
<i>“Smart city as a high-tech intensive and advanced city that connects people, information and city elements using new technologies in order to create a sustainable, greener city, competitive and innovative commerce, and an increased life quality.”(KOMINOS, 2011)</i>
<i>“Smart city [refers to] a local entity - a district, city, region or small country -which takes a holistic approach to employ[ing] information technologies with real-time analysis that encourages sustainable economic development.” (IDA, 2012)</i>
<i>“The application of information and communications technology (ICT) with their effects on humancapital/education,social and relational capital, and environmental issues is often indicated by the notion of smart city.” (LOMBARDI et al., 2012)</i>
<i>“A smart city infuses information into its physical infrastructure to improve conveniences, facilitate mobility, add efficiencies, conserve energy, improve the quality of air and water, identify problems and fix them quickly, recover rapidly from disasters, collect data to make better decisions, deploy resources effectively, and share data to enable collaboration across entities and domains.” (NAM; PARDO, 2011)</i>
<i>“Creative or smart city experiments [...] aimed at nurturing a creative economy through investment in quality of life which in turn attracts knowledge workers to live and work in smart cities. The nexus of competitive advantagehas[...]shifted to those regions that can generate, retain, and attract the best talent.” (THITE, 2011)</i>
<i>“A smart city is understood as a certain intellectual ability that addresses several innovative socio-technical and socio-economic aspects of growth. These aspects lead to smart city conceptions as “green” referring to urban infrastructure for environment protection and reduction of CO2 emission, “interconnected” related to revolution of broadband economy, “intelligent” declaring the capacity to produce added value information from the processing of city’s real-time data from sensors and activators, where as the terms“innovating”,“knowledge”cities interchangeably refer to the city’s ability to raise innovation based on knowledgeable and</i>

Conceito
<i>creative human capital.</i> ”(ZYGIARIS, 2013)
<i>“Smart Cities initiatives try to improve urban performance by using data, information and information technologies (IT) to provide more efficient services to citizens, to monitor and optimize existing infrastructure, to increase collaboration among different economic actors, and to encourage innovative business models in both the private and public sectors.” (MARSAL-LLACUNA et al., 2014)</i>
<i>“Being a smart city means using all available technology and resources in an intelligent and coordinated manner to develop urban centers that are at once integrated, habitable, and sustainable.” (BARRIONUEVO et al., 2012)</i>
<i>“A city is smart when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance.” (CARAGLIU et al., 2011)</i>
<i>“Smart cities will take advantage of communications and sensor capabilities sewn into the cities’ infrastructures to optimize electrical, transportation, and other logistical operations supporting daily life, thereby improving the quality of life for everyone .”(CHEN, 2010)</i>
<i>“Smart community – a community which makes a conscious decision to aggressively deploy technology as a catalyst to solving its social and business needs – will undoubtedly focus on building its high-speed broadband infrastructures, but the real opportunity is in rebuilding and renewing a sense of place, and in the process a sense of civic pride. [...] Smart communities are not, at their core, exercises in the deployment and use of technology, but in the promotion of economic development, job growth, and an increased quality of life. In other words, technological propagation of smart communities isn’t an end in itself, but only a means to reinventing cities for a new economy and society with clear and compelling community benefit.” (EGER, 2009)</i>
<i>“A smart city is based on intelligent exchanges of information that flow between its many different subsystems. This flow of information is analyzed and translated into citizen and commercial services. The city will act on this information flow to make its wider ecosystem more resource efficient and sustainable.” (GARTNER, 2011).</i>
<i>“A city well performing in a forward-looking way in economy, people,</i>

Conceito
<i>governance, mobility, environment, and living, built on the smart combination of endowments and activities of self-decisive, independent and aware citizens. Smart city generally refers to the search and identification of intelligent solutions which allow modern cities to enhance the quality of the services provided to citizens.” (GIFFINGER et al., 2007)</i>
<i>“A smart city, according to ICLEI, is a city that is prepared to provide conditions for a healthy and happy community under the challenging conditions that global, environmental, economic and social trends may bring.”(GUAN, 2012)</i>
<i>“A city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, seaports, communications, water, power, even major buildings, can better optimize its resources, plan its preventive maintenance activities, and monitor security aspects while maximizing services to its citizens.”(HALL, 2000)</i>
<i>“A city connecting the physical infrastructure, the IT infrastructure, the social infrastructure, and the business infrastructure to leverage the collective intelligence of the city.” (HARRISON et al. 2010)</i>
<i>(Smart) cities as territories with high capacity for learning and innovation, which is built-in the creativity of their population, their institutions of knowledge creation, and their digital infrastructure for communication and knowledge management.” (KOMNINOS, 2011)</i>

Fonte: Adaptado de ALBINO, BERNARDI e DANGELICO (2015).

Em seguida, é apresentado o tema *big data*, que possui forte associação com o conceito de *smart cities*, conforme segue.

2.2 BIG DATA

No ano de 2011, o número médio diário de postagens (*tweets*) publicados no site Twitter foi de 200 milhões (TWITTER, 2011); no site Facebook.com, 30 bilhões de conteúdos são compartilhados mensalmente (MANYIKA et al., 2011); a cada minuto são postados 72 horas de vídeo no Youtube (WOOLLASTON, 2013). Outro exemplo é a rede Walmart, que lida a cada hora com 1 milhão de transações em um banco de dados de aproximadamente 2,5 petabytes (10^{15} bytes) (ALLIANCE, 2014), o que corresponde aproximadamente a 170 vezes a quantidade de dados da Biblioteca do Congresso Americano (que possui

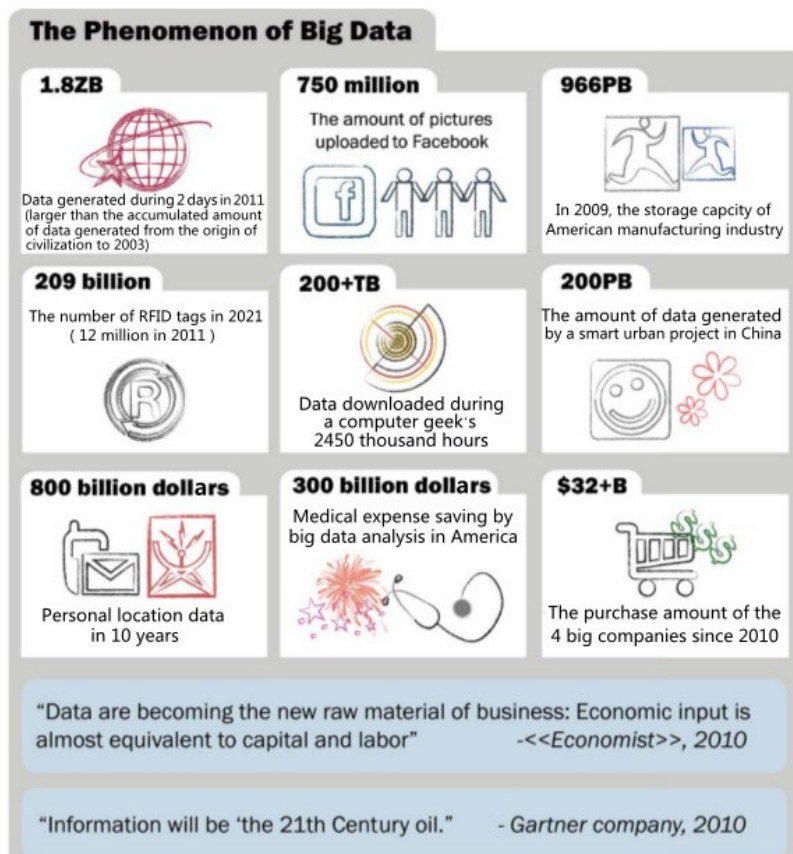
em torno de 36,8 milhões de livros catalogados) (BETTINO, 2012). Segundo a IBM (2013), 90% dos dados existentes no mundo atual foram criados nos últimos dois anos. E ainda, em 2020, a tendência é que aproximadamente 26 bilhões de aparelhos sejam conectados a IoT (internet das coisas, do inglês, *internet of things*) (RIVERA; MEULEN, 2013). Na Figura 3 este cenário pode ser melhor visualizado.

Toda esta quantidade de dados produzida corresponde ao fenômeno conhecido como *big data*. Este tema tem sido comumente relacionado a volumes de dados extremamente grandes, no entanto, existem diversas definições na literatura que evidenciam outras de suas características (CHEN; MAO; LIU, 2014). Fan e Bifet (2012), citando Laney (2001), definem o chamado modelo 3Vs:

- *Volume*: há hoje mais dados do que jamais houve (quantidade), porém, não há tantas ferramentas que possam processá-los;
- *Variety*: existem muitos tipos diferentes de dados, como texto, dados de sensores, audio, vídeos, imagens etc;
- *Velocity*: os dados chegam a fluxos contínuos, e necessita-se analisá-los em tempo real.

Nos anos seguintes a este modelo, segundo Chen, Mao e Liu (2014), citando Zikopoulos et al. (2011) e Meijer (2011), empresas como a Gartner, Microsoft e IBM corroboraram e adaptaram este conceito em seus departamentos de pesquisa, com as seguintes definições: *volume*, corresponde a grandes massas de dados e altas escalas; *velocity*, refere-se a linha de tempo dos dados, que deve ser coletado e analisado rapidamente, para manter o valor comercial; e *variety*, que significa os vários formatos de dados, incluindo dados estruturados, semi e não estruturados. Seguindo esta linha, existem várias definições no meio acadêmico e de negócios.

Figura 3: O fenômeno *big data*



Fonte: CHEN; MAO; LIU (2014).

Abaixo, segue uma lista de algumas das definições encontradas na literatura, sob diferentes perspectivas e aspectos, que contribuem no entendimento deste fenômeno:

- *Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast, or doesn't fit the structures of your database architectures.* (O'REILLY, 2012);
- *[...] a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of*

data, by enabling the high-velocity capture, discovery, and/or analysis. (GANTZ; REINSEL, 2011);

- Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. (GARTNER, [20--]);

- Big Data refers to data that is too big to fit on a single server, too unstructured to fit into a row-and-column database, or too continuously flowing to fit into a static data warehouse. While its size receives all the attention, the most difficult part of big data really involves its lack of structure. (DAVENPORT, 2014);

- Big Data is a new term used to identify datasets that we can not manage with current methodologies or data mining software tools due to their large size and complexity. (FAN; BIFET, 2012).

Ainda definindo este conceito, alguns autores adicionam novas características, como *veracity*, que diz respeito ao fato de que estes dados devem ser confiáveis para que as organizações possam tomar suas decisões baseadas neles (IBM, 2013). Segundo Fan e Bifet (2012, tradução nossa), hoje em dia também são consideradas duas dimensões importantes em *big data*:

- *Variability*: corresponde aos diferentes esquemas dos dados coletados e como estes são interpretados;
- *Value*: valor de negócio, que possibilita às organizações tomarem decisões que antes (de *big data*) não era possível.

Como causas deste fenômeno pode-se apontar o avanço do uso de TICs em diversos setores da sociedade (GROBELNIK, 2012). O uso de *cloud computing* e da *web 2.0*, assim como a alta inserção de *smartphones*, sensores e outros dispositivos produtores de dados na rotina das pessoas, são em termos gerais, características do avanço tecnológico que vivemos nos dias atuais. A era da informação e estes avanços com certeza contribuem para um maior volume, velocidade e variedade de dados produzidos. Grobelnik (2012), citando Manyika et al. (2011) e Hillbert e Lopez (2011), destaca os principais fatores de causa do fenômeno *big data*:

- O aumento mundial da capacidade de armazenamento: considerando os aparelhos digitais e analógicos produzidos no período entre o ano de 2000 e 2007, estima-se que houve um crescimento de aproximadamente 50 exabytes (1 bilhão de gigabytes) para em torno de 290 exabytes;
- O aumento mundial da capacidade de processamento de dados: entre 2000 e 2007, esta capacidade aumentou de 10^{12} milhões de instruções por segundo para $6,5 \times 10^{12}$. Os autores explicam que estes valores foram estimados através da multiplicação da quantidade de aparelhos instalados pela sua capacidade de processamento;
- A alta disponibilidade de dados: o mundo e as organizações possuem muito mais dados disponíveis. Segundo os autores, somente nos EUA, organizações de diversos setores, com mais de 1.000 empregados, possuem no mínimo 100 TB de dados armazenados em média. Além disso, o aumento do uso das redes sociais e da internet das coisas contribuiu significativamente para o aumento desta disponibilidade.

Em seguida, é detalhada uma das fontes de dados do fenômeno *big data*, conhecida como *open data* - dados abertos.

2.3 OPEN DATA

Conforme afirmado anteriormente, governos de diversos países, como o dos EUA (U.S. GOVERNMENT, 2015), do Reino Unido (REINO UNIDO, 2015), Brasil (BRASIL, 2015) e diversos outros, além de setores da iniciativa privada e do meio acadêmico tem investido na abordagem Open Data (dados abertos) (MANYKA et al., 2013). É possível ter uma ideia melhor sobre o tamanho deste movimento na medida em que se ilustra esta tendência em forma de números. No ano de 2012, a Web Foundation publicou um ranking chamado de Open Data Index, onde mais de 40 países são classificados conforme o nível de uso de open data, cujo resultado provém de um conjunto de indicadores relacionadas ao uso de dados abertos (ALONSO, 2012), conforme Quadro 2 e ilustrados na Figura 4. Este ranking foi derivado do Web Index, cuja organização mantenedora é a própria Web Foundation, fundada por Tim Berners-Lee, que tem como foco medir o impacto social do uso da internet no mundo (EDWARDS, 2012).

Ainda trazendo mais informações sobre este contexto, conforme relatório da empresa McKinsey, existem atualmente mais de 1 milhão de datasets (bases de dados) abertas (open data) em governos no mundo

inteiro (levantamento no ano de 2013), e que representam uma estimativa de geração de mais de 3 trilhões de dólares na economia mundial, dividindo-se em aproximadamente \$1,1 trilhão nos EUA , \$900 trilhões na Europa e \$1,7 trilhões no resto do mundo (MANYKA et al., 2013). Estes montantes estimados se baseiam no uso de dados abertos governamentais pelo governo e sociedade, que devem melhorar a eficiência de processos organizacionais destes setores (públicos e iniciativa privada), criação de novos produtos e serviços, além de benefícios sociais não quantificados (MANYKA et al., 2013, tradução nossa). Para se ter uma ideia, a Climate Corporation, uma startup que combina 30 anos de dados meteorológicos e 14 terabytes de dados de solo, oriundos de agências governamentais, foi recentemente adquirida por cerca de US\$ 1 bilhão (MANYKA et al., 2013; KESMODEL, 2013, tradução nossa).

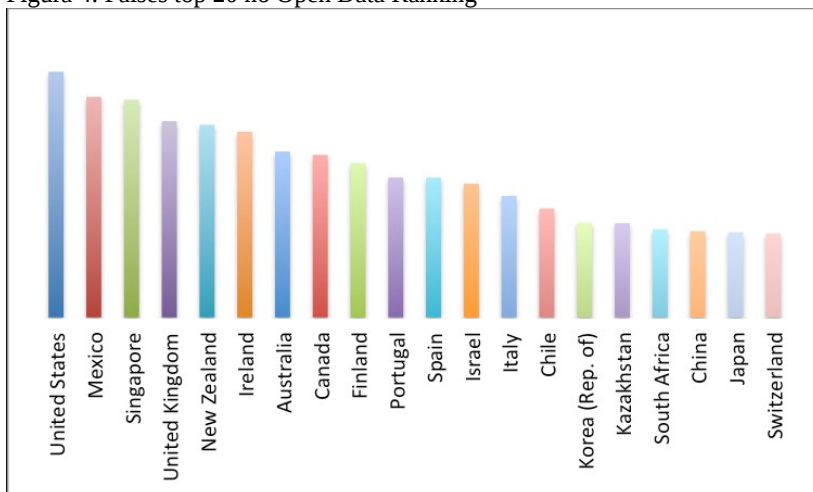
Quadro 2: Amostra sobre a composição do *Open Data Index*

Indicador	Questão para construção
<i>Government use of open licenses</i>	<i>Survey Question: To what extent are government agencies publishing information on the web using open licenses? [1 = not at all; 10 = extensively]</i>
<i>Publication of trade data on the Web</i>	<i>Survey Question: To what extent are there government data on the web for international trade? [1 = no data at all; 10 = extensive data]</i>
<i>Publication of health data on the Web</i>	<i>Survey Question: To what extent are there government data on the web on health sector performance (hospitals, doctors etc)? [1 = no data at all; 10 = extensive data]</i>
<i>Publication of education data on the Web</i>	<i>Survey Question: To what extent are there government data on the web on educational performance? [1 = no data at all; 10 = extensive data]</i>
<i>Publication of transport data on the Web</i>	<i>Survey Question: To what extent are there government data on the web on transport data and schedules? [1 = no data at all; 10 = extensive data]</i>
<i>Publication of census data on the Web</i>	<i>Survey Question: To what extent are there government data on the web for the census - age, income, voting, migration etc? [1 = no data at all; 10</i>

Indicador	Questão para construção
	<i>= extensive data]</i>
<i>Publication of map data on the Web</i>	<i>Survey Question: To what extent are there government map data on the web? [1 = no data at all; 10 = extensive data]</i>
<i>Tax filing via the Web</i>	<i>Survey Question: To what extent are there government data on tax returns and how to submit these? [1 = no data at all; 10 = extensive data]</i>
<i>Information on contacts in government departments</i>	<i>Survey Question: To what extent are there government data on the web on whom to reach for different government services (e.g. local police stations, libraries etc)? [1 = no data at all; 10 = extensive data]</i>
<i>Publication of crime data on the Web</i>	<i>Survey Question: To what extent are there government data on the web on crime in your country? [1 = no data at all; 10 = extensive data]</i>
<i>Ease of access of government data</i>	<i>Survey Question: How easy is it easy to access government data (as listed in Q23) on the Web in open machine readable format (.csv or .xls files, XLS, RDF etc)? [1 = very difficult to access government data; 10 = very easy to access government data]</i>
<i>Extent of Open Government Data Initiative</i>	<i>Survey Question: Does the government have a specific open data initiative? [1 = no full open government data initiative; 10 = full open government data initiative]</i>
<i>Creation of new services based on government data</i>	<i>Survey Question: To what extent are Web applications and services in areas such as health, education, security, budgets etc "built" on top of government data (i.e. there has been new and useful information and services derived from the published government data in those fields? [1 = none at all; 10 = extensive applications and services]</i>

Fonte:O autor. Dados extraídos de *WebFoundation* (2012).

Figura 4: Países top 20 no Open Data Ranking



Fonte: Alonso(2012).

Quando se estuda dados abertos, muito se encontra sobre dados abertos governamentais e *open government*. É interessante portanto fazer esta distinção. Dados Abertos (ou *Open Data*) são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa, sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras (OPENDATAHANDBOOK, 2014). Tornar um dado aberto significa deixá-lo sob uma licença que não restrinja o seu uso e compartilhamento por terceiros, indiferentemente de qual seja a finalidade do uso, e ainda disponibilizá-lo se possível via internet, conforme segue:

- Reutilização e Redistribuição: os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados;
- Disponibilidade e Acesso: os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, preferencialmente possíveis de serem baixados pela internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável;

- Participação Universal: todos devem ser capazes de usar, reutilizar e redistribuir - não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso 'não-comercial' que impediriam o uso 'comercial', ou restrições de uso para certos fins (ex.: somente educativos) excluem determinados dados do conceito de 'abertos' (OPENDATAHANDBOOK, 2014).

Similarmente, o termo governo aberto (open government) está relacionado ao conceito de governo eletrônico. Segundo Governoeletrônico (2005), o e-gov (Governo Eletrônico) pode ser entendido como uma das principais formas de modernização do Estado e está fortemente apoiado no uso das novas tecnologias para a prestação de serviços públicos, mudando a maneira com que o governo interage com os cidadãos, empresas e outros governos. Rover (2008) afirma que o governo eletrônico tem se constituído em uma infra-estrutura de rede compartilhada por diferentes órgãos públicos a partir da qual a gestão dos serviços públicos é realizada. Chahin et al. (2004) afirmam que o Governo Eletrônico possui o objetivo de estabelecer um novo paradigma cultural de inclusão digital centrado no cidadão, propiciando também uma redução de custos, melhoria na gestão e qualidade dos serviços públicos, uma maior transparência e simplificação de processos. Desde sua criação, este conceito levou a diversas iniciativas de aplicar TICs para melhorar os processos operacionais e administrativos públicos, incluindo desde a melhoria dos serviços prestados aos cidadãos até a iniciativas voltadas para a disponibilização de dados governamentais abertos (RIBEIRO; PEREIRA, 2015).

Finalmente, unindo dados abertos (open data) e governo aberto (open government), pode-se entender melhor o termo Dados Governamentais Abertos (DGA), que trata da disponibilização através da Internet de informações e dados governamentais de domínio público para a livre utilização pela sociedade (AGUNE; GREGORIO FILHO; BOLLIGER, 2010). Conforme o Laboratório Brasileiro de Cultura Digital (2011), os principais benefícios da abertura de dados governamentais são: transparência e controle democrático; participação popular; empoderamento dos cidadãos; melhores ou novos produtos e serviços privados; inovação; melhora na eficiência de serviços governamentais; melhora na efetividade de serviços governamentais; medição do impacto das políticas; conhecimento novo a partir da combinação de fontes. Segundo o Tribunal de Contas da União

(BRASIL, 2015), corroborando estas justificativas, pode-se resumir em cinco os principais motivos: transparência na gestão pública; contribuição da sociedade com serviços inovadores ao cidadão; aprimoramento na qualidade dos dados governamentais; viabilização de novos negócios; obrigatoriedade por lei (no caso do Brasil). Além disso, para Diniz (2010), a disponibilização de dados governamentais abertos tem o potencial de promover a inclusão digital, se fornecidos em formatos que respeitem os padrões de acessibilidade (para pessoas com deficiência física), e ainda promover maior entendimento sobre as ações do governo. O Brasil, para esse contexto em nosso território, vem aderindo a este movimento. A Lei de acesso à informação (BRASIL, 2011) garante o direito constitucional de acesso a informações produzidas ou detidas pelo Governo. Conforme Matheus, Vaz e Ribeiro (2014), apud Vaz et al. (2011), pode-se identificar três níveis de abertura de dados no Brasil desde o crescimento deste movimento:

Fase 1 – Páginas Estáticas com Relatórios Financeiros: Websites governamentais com informações, geralmente das finanças públicas, onde podem ser visualizadas parte destas informações em formato estático;

Fase 2 – Portais de Transparência: Portais mais robustos com bases de dados maiores e com possibilidade de buscas. Ainda não era possível estabelecer cruzamentos entre os dados;

Fase 3 – Dados Governamentais Abertos (DGA): Portais com possibilidade de realizar consultas com sistema de filtragem e de grande número de datasets em formato aberto. O formato aberto facilita o cruzamentos das informações em busca de resultados específicos sobre o governo (MATHEUS; VAZ; RIBEIRO, 2014).

Conforme Shadbolt et al. (2012), o potencial do uso de open data ainda pode ser aumentado se forem utilizadas técnicas e padrões da web semântica, como linked data (dados ligados) e RDF (Resource Description Framework) por exemplo, para interligar estes dados com outras bases disponíveis, através de ligações (links) semânticas (passíveis de interpretação por computadores e não apenas por humanos) (BERNERS-LEE, 2006). Esta combinação de Linked Data Web e Open Government, como descrito no trabalho Linked Open Government Data, já pode ser vista em diversos projetos pelo mundo (SHADBOLT et al., 2012).

2.4 GOVERNANÇA DE DADOS

Ao trabalhar com dados de diversas fontes distintas é necessário conhecer o conceito Governança de Dados. Governança de dados é uma disciplina da área de controle de qualidade que adiciona um novo rigor ao processo de gestão, utilização, melhoria e proteção de informação organizacional (IBM, 2007, tradução nossa). Conforme Panian (2010), a governança de dados pode ser compreendida como o conjunto de processos, políticas, padrões, estruturas e tecnologias para gerenciar e assegurar disponibilidade, acessibilidade, qualidade, consistência, auditabilidade e segurança nos dados da organização. É possível percebê-la como um instrumento para se implementar diretrizes para controlar os dados de uma determinada organização. Segundo Barata (2015), a GD (governança de dados) tem como responsabilidade o controle e gerenciamento dos dados da organização, sendo ela a responsável por todo o controle dos seus dados e a sua transformação em informação organizacional.

Segundo Mosley et al. (2009), a gestão de dados visa o planejamento, execução e fiscalização de políticas, práticas e projetos para adquirir, controlar, proteger e alavancar eficazmente o uso dos ativos de dados na organização. Dentre algumas metodologias existentes na literatura se implementar a GD nas organizações, pode-se citar os programas de Governança de Dados da Oracle, da IBM, e do DGI (Data Governance Institute) (SANTOS, 2010), além do DMBOK (Data Management Body of Knowledge) (DAMA, 2015).

Conforme Barata (2015), o foco de atuação da área de GD pode variar de organização para organização, começando por exemplo desde privacidade até a qualidade dos dados, porém, possuindo sempre objetivos comuns a serem atingidos. Fernandes e Abreu (2012) listam objetivos em comum em qualquer política de GD, conforme destacados abaixo:

- Permitir uma melhor tomada de decisões;
- Proteger as necessidades dos *stakeholders*;
- Institucionalizar uma gerência comum no tratamento de problemas de dados;
- Construir padrões, processos e metodologias que possam ser disseminadas pela organização;
- Reduzir custos e aumentar a eficácia através da coordenação de esforços conjuntos;
- Garantir a transparência dos processos (FERNANDES;ABREU,2012).

Conforme Pacheco et al. (2012), a GD pode contribuir muito para a área de governo eletrônico (e-gov). Projetos de governo aberto lidam com dados de diversas agências governamentais e a GD pode auxiliar a criar diretrizes em comum para essa integração.

2.4.1 Qualidade De Dados

A Qualidade de Dados é uma das áreas de conhecimento abordadas pela Governança de Dados na estrutura fornecida pelo DMBOK (*Data Management Body of Knowledge*), desenvolvido pela organização internacional sem fins lucrativos DAMA (*Data Management Association International*), conforme pode ser visto na Figura 5 (BARBIERI, 2013; DAMA, 2015). Conforme Britto e Almeida Júnior (2006), define-se Qualidade de Dados como o grau de aderência entre as visões apresentadas pelos dados armazenados e os mesmos dados no mundo real. Ou seja, considerando o dado como uma representação de algum elemento do mundo real, quanto melhor for essa representação (mais próxima da realidade), melhor será a qualidade do dado. De acordo com Wang (1996, traduzido e adaptado), a qualidade de dados é definida na literatura como um conceito multidimensional, cujas dimensões mais citadas são: precisão, completude, consistência, e frequente atualização.

Figura 5: Governança de Dados e Qualidade de Dados segundo DMBOK



Fonte: DAMA(2015).

Conforme Salvador et al. (2006), um dado tem qualidade se satisfaz os seguintes requisitos de uso:

- **Preciso:** é a medida de quão correto, quão livre de erros, quão próximo está este dado do fato verdadeiro. É medida fundamental da qualidade de dados; se um dado não é correto, as outras dimensões são menos importantes. Para ser correto, um valor deve ser certo e deve ser representado de uma forma consistente e sem ambiguidade.
- **Disponível em tempo:** os dados estão suficientemente atualizados para as tarefas que os necessitam;
- **Relevante:** importante para o tomador de decisões em um contexto; é útil e aplicável à tarefa em questão;

- **Completo:** deve conter todos os fatos importantes, na amplitude e profundidade adequadas às suas necessidades;
- **Simples/Compreendido:** evitando a chamada “sobrecarga de informação”;
- **Confiável:** depende da fonte ou método de coleta. (SALVADOR et al., 2006).

Ainda conforme Salvador et al. (2006, apud Wang, 1998), outras dimensões podem ser encontradas revisando a literatura, são elas: livre de erro, objetividade, reputação, credibilidade, relevância, valor adicionado, em tempo, quantidade apropriada de informação, capacidade de interpretação, compreensão, representação consistente, representação concisa, facilidade de manipulação, acessibilidade e segurança.

Além das dimensões de Qualidade de Dados encontradas na literatura, o conceito de Proveniência de Dados também está associado a esta área. Conforme Melo et al. (2011), apud Veregin (1995), a proveniência (dos dados) pode ser uma métrica de qualidade importante, já que o processo de derivação dos dados tem implicações significativas na qualidade de dados e nos erros introduzidos por dados falhos, que aumentam na medida em que estes se propagam. Conforme Buneman, Khana e Tan (2001), proveniência de dados (também conhecida como linhagem ou *pedigree* de dados) é definida como a descrição da origem de um pedaço (*stream*) de dado e o processo pelo qual este chegou ao banco de dados. A proveniência dos dados pode ser entendida como metadados que descrevem a origem de um dado e como este foi capturado. Conforme Marins (2008), este conceito é utilizado com padrões de metadados como o Dublin Core e o Warwick Framework, além da norma ISO 19115:2003, podendo ser entendido como informação sobre a origem de um item de dado. Recentemente, modelos como o *Open Provenance Model* e o PROV (do W3C) foram desenvolvidos para representar a origem de dados, como respostas às dificuldades de identificar a origem de dados de experimentos científicos (BIVAR et al., 2013).

2.4.2 Qualidade Da Informação

A Qualidade da Informação em geral é tratada sem distinção em relação à Qualidade de Dados na literatura. Conforme Santos (2009), o termo Qualidade da Informação (QI), fortemente em destaque em pesquisas a partir dos anos 90, é utilizado na literatura como sinônimo

de Qualidade de dados, considerando em geral o dado como uma informação bruta nestas pesquisas. Bettancourt (2015 apud Pipino; Lee; Yang, 2002) também afirma que a Qualidade de Dados é tratada como sinônimo de Qualidade da Informação. Para os autores, a qualidade da informação está fortemente atrelada à qualidade dos dados, possuindo praticamente as mesmas dimensões de qualidade, com apenas algumas variantes, conforme pode ser visto no Quadro 3.

Quadro 3: Dimensões de qualidade da informação

Categoria	Dimensão	Definição
Intrínseca	Credibilidade	Quão os dados armazenados são confiáveis e críveis?
Intrínseca	Livre de erros	Quão os dados estão corretos e livres de erros?
Intrínseca	Objetividade	Quão os dados são não viesados, imparciais e não apresentam nenhum tipo de prejuízo?
Intrínseca	Reputação	Quão os dados são confiáveis com relação as suas origens?
Representacional	Representação concisa	Quão os dados são representados compactamente?
Representacional	Representação consistente	Quão os dados são representados nos mesmos formatos?
Representacional	Interpretabilidade	Quão os dados estão representados em linguagem apropriada, utilizando símbolos e códigos corretos e apresentam definições claras e precisas?
Representacional	Compreensividade	Quão fácil é compreender os dados exigidos para uma determinada tarefa?
Acessibilidade	Acessibilidade	Quão os dados estão disponíveis ou fácil e rapidamente recuperáveis?
Acessibilidade	Facilidade de manipulação	Quão os dados são fáceis de manipular e de aplicar em tarefas diversas?

Categoria	Dimensão	Definição
Contextual	Quantidade apropriada de dados	Quão o volume de dados é apropriado para a tarefa a que se destinam?
Contextual	Compleitude	Quão os dados são não ausentes e suficientes (ampla e profundamente) para a tarefa que se destinam?
Contextual	Relevância	Quão os dados são aplicáveis e úteis para a tarefa a que se destinam?
Contextual	Tempestividade	Quão os dados são atualizados para a tarefa a que se destinam?
Contextual	Valor agregado	Quais os benefícios e vantagens que trazem os dados para o sistema que deles se utilizam?

Fonte: Bettancourt (2015), citando Strong, Lee e Wang (1997); Pipino; Lee; Yang (2002).

Corroborando esta ideia, conforme Assis e Moura (2011), citando Knight e Burn (2005), em muitas pesquisas dedicadas ao contexto dos sistemas de informação, a Qualidade da Informação é um termo substituível por Qualidade de Dados. Para Assis e Moura (2011), na literatura sobre Qualidade da Informação encontram-se duas abordagens distintas: positivista, que utiliza bases empíricas e quantitativas para a elaboração das noções de qualidade; e pragmática, que considera que a Qualidade da Informação é essencialmente subjetiva e como tal, não é passível de ser operacionalizada.

2.5 MODELOS CONCEITUAIS

Um modelo é uma representação de algo, ou “... uma abstração de uma realidade de acordo com uma conceitualização” (CASTRO; BAIÃO; GUIZZARDI, 2009 apud GUIZZARDI, 2005, traduzido e adaptado). Modelos conceituais, ou modelagem de conceitos, são utilizados em diversas pesquisas e disciplinas, com seu foco variando conforme a finalidade de seu uso. Em geral, pesquisadores utilizam modelos conceituais para estruturar e guiar suas pesquisas (PREEZ; MEYER, 2011). Diversos trabalhos são encontrados na literatura que

utilizam modelos conceituais aplicados nas mais distintas áreas, como na área da saúde por exemplo.

Dentre as áreas que usam mais frequentemente modelos conceituais, duas se destacam: a Engenharia de Software (mais precisamente, projetos de Banco de Dados), e a Engenharia de Conhecimento. Na Engenharia de Software, o modelo conceitual de uma aplicação é utilizado para definir os conceitos que envolvem uma aplicação, suas funcionalidades e possíveis interações (TELECKEN, 2004, apud SOUZA et al., 1999 e HEUSER, 1991). Ao desenvolver projetos de Banco de Dados por exemplo, um modelo conceitual representa fatos em alto nível de abstração em um determinado domínio a ser modelado. Conforme Navathe (1992), o modelo conceitual em banco de dados é uma etapa da modelagem de dados em projetos de desenvolvimento de Bancos de Dados, e captura os conceitos essenciais no seu ambiente (domínio). Na Engenharia do Conhecimento, diferentemente de projetos de Banco de Dados, conceitos são modelados geralmente com o uso de ontologias, com o objetivo de representar conhecimento. A diferença entre ontologias e modelos de dados em banco de dados está essencialmente em suas distintas finalidades, conforme segue:

Um modelo de dados é um modelo de informação em um restrito e bem delimitado domínio de aplicação, enquanto que uma ontologia é destinada a fornecer um conjunto de conceitos compartilhados por vários usuários e aplicações. [...] um modelos de dados vive em um mundo relativamente pequeno e fechado; ontologias são destinadas para um mundo aberto e distribuído (SCHREIBER, 2008, tradução nossa).

Modelos conceituais auxiliam a compreender uma realidade, geralmente modelando conceitos, suas definições e relações chaves para um determinado domínio de estudo. Definindo de uma forma mais genérica, conforme Fawcett (2014), tradução nossa, um modelo conceitual se refere a um conjunto de conceitos (e as relações entre estes) relativamente abstratos e genéricos, que identificam fenômenos de interesse de estudo para uma determinada disciplina. A essência de modelagem de conceitos está em abstrair os principais conceitos, defini-los, e identificar suas relações. Bunge (2003) contribui para modelar estes itens na medida em que fornece um método que permite visualizar de forma sistêmica um determinado objeto de estudo. O autor fornece a

quádrupla CESM (Componente, Ambiente, Estrutura e Mecanismos), que permite modelar sistemas de diversas naturezas (ex.: sistemas de trânsito, sistemas biológicos etc). Com o método CESM, é possível identificar em um determinado sistema, seus principais conceitos, atores, suas relações entre si e entre seu meio de atuação, bem como também, modelar os seus principais processos (mecanismos). Isso auxilia na modelagem de conceitos chaves da estrutura e do funcionamento de qualquer objeto de estudo que possa ser pensado como um sistema.

2.5.1 Etapas De Construção De Modelos Conceituais

Conforme Monteiro (2006), para fundamentar a definição de conceitos, parte fundamental da modelagem conceitual, a Teoria do Conceito de Dahlberg (1978) e a Teoria da Classificação Facetada, de Raganathan (1967) devem ser estudadas.

Monteiro (2006), citando Dahlberg (1978), afirma que a Teoria do Conceito estrutura a formação de conceitos na medida em que estes são definidos e são estabelecidas suas composições e relacionamentos. Nesta teoria, além da definição dos conceitos, suas partes, relacionamentos e hierarquias, é buscado identificar a função de cada conceito, ou seja, qual o papel de cada conceito em um determinado domínio de estudo.

Em relação à Teoria de Classificação Facetada de Raganathan (1967), Monteiro (2006) afirma que a estrutura de um determinado domínio deve ser entendida com a exteriorização do conhecimento organizado conforme inferências e abstrações pretendidas. Para isso, a Teoria de Raganathan auxilia ao propor um conjunto de categorias para identificar conceitos hierárquicos, de acordo com sua natureza, que são: Matéria (característica sendo descrita), Espaço (aspecto de localização), Tempo (aspecto temporal) e Personalidade (entidades relacionadas), definindo não apenas relações hierárquicas, mas relações do tipo parte-todo (MONTEIRO, 2006).

No âmbito de projetos de desenvolvimento de Banco de Dados e Engenharia de Software, modela-se conceitos para que estes organizem os dados a serem armazenados e manipulados pelos sistemas de informação. Conforme Batini, Cere e Natathe (1992), um modelo de dados é uma coleção de conceitos que podem ser usados para descrever um conjunto de dados e as operações para manipular esses dados. A modelagem conceitual é uma das primeiras etapas em projetos de desenvolvimento de sistemas de informação. Para a construção de modelos conceituais, utiliza-se em geral o MER – Modelo Entidade

Relacionamento, que é uma representação da realidade através de entidades, seus relacionamentos e atributos (LONDEIX, 1995). Para a construção destes modelos, são utilizadas as consolidadas linguagens de notação visual para a construção de diagramas. Uma das notações mais utilizadas e citadas na literatura é a de Peter Chen, de 1976. Além de diagramas ER (Entidade-Relacionamento), a linguagem UML também pode ser utilizada para modelar entidades e suas relações. Estes artefatos também são chamados de Ontologias de Software, pois modelam esquemas conceituais, com o foco principal nas atividades de armazenamento e manipulação de dados, construindo a consistência dos dados de um sistema de informação (ROUSSEY et al., 2011, tradução nossa).

A Engenharia de Conhecimento também auxilia na modelagem de conceitos. Esta disciplina foca em teorias, métodos e ferramentas para desenvolver aplicações intensivas em conhecimento (SCHREIBER, 2008), provendo um arcabouço para realizar essa atividade. Mais precisamente, a disciplina de Engenharia de Ontologias fornece norteamentos para construir conceitualizações de diversos domínios. As chamadas Ontologias de Aplicação, Ontologias de Domínio, Ontologias de Referência de Núcleo, Ontologias Genéricas, e Ontologias de Alto Nível são utilizadas para modelagem de conceitos, com o objetivo de compartilhar conceitualizações em qualquer domínio de conhecimento (ROUSSEY et al., 2011). Elas distinguem-se entre si por possuírem diferentes níveis de granularidade. Conforme Roussey et al., (2011), elas começam representando pequenos universos de conhecimento e vão aumentando seu escopo, representando maiores áreas de conhecimento, tornando-se mais genéricas (iniciando em Ontologias de Domínio e terminando em Ontologias de Alto Nível). Para construí-las, são utilizados métodos como por exemplo o Methontology (GOMEZ-PEREZ et al., 2004), On-to-Knowledge (FENSEL & HERMELEN, 2008), o guia 101 (NOY & MCGUINNESS, 2008) e o OntoKEM (RAUTENBERG et al., 2009). Como linguagens muitas vezes utilizadas por estes métodos, dentre várias destacam-se: *Description Logics* (DL), *Conceptual Graphs* (CG), *First Order Logic* (FOL) (SCHREIBER, 2008), OntoUML (GUIZZARDI, 2005) e a OWL (*Web Ontology Language*). As ferramentas Protégé e OntoKEM também são exemplos que podem ser utilizados como complementos a estes métodos.

2.6 TRABALHOS RELACIONADOS

Implantar uma *smart city* (implementar seus conceitos) e ainda

trabalhar com *big data* são atividades que exigem grandes desafios. Vilajosana et al. (2013) propõem um modelo para se implementar uma *smart city* através do uso de *big data*. Neste trabalho, primeiramente os autores explicam que as grandes cidades atuais possuem necessidades reais de se tornarem *smart* devido a três fatores principais: mais da metade da população vive em ambientes urbanos (nas cidades europeias, 80% inclusive); os esforços para facilitar as condições de vida (ex.: mobilidade urbana) nestes ambientes são enormes; e os efeitos da urbanização no clima global vêm se evidenciando cada vez mais. Eles afirmam que, observando estes fatores, governos tem procurado por soluções tecnológicas para estes problemas o que consequentemente disparou o interesse de grandes fornecedores de serviços e produtos, como a IBM, Cisco, HP dentre outros, a produzirem suas próprias soluções. Porém, mesmo com ambas as partes interessadas (governo e indústria), os investimentos em grande escala, principalmente na Europa, não têm ocorrido e, quando ocorrem, têm encontrado diversas barreiras. Os autores então apontam os três principais desafios que impedem este desenvolvimento:

- **Desafio Político/ Administrativo:** existe uma necessidade de mudanças nas estruturas político/administrativas das cidades. Por exemplo, uma prefeitura geralmente possui diversos setores e departamentos que lidam com as diversas áreas que uma *smart city* deve se preocupar (meio ambiente, tecnologias, turismo etc). No entanto, estes setores dificilmente funcionam de maneira integrada, principalmente do ponto de vista dos seus sistemas de informação, mas também nos processos de tomadas de decisões;
- **Desafio Tecnológico:** há certa complexidade de desenvolvimento nas soluções tecnológicas voltadas para as *smart cities/big data*. Estas soluções devem lidar com diversos tipos de dados, de diversas fontes, em diversos modelos e formatos e, em altos volumes e velocidade de produção (como visto na definição de *big data*);
- **Desafio financeiro:** é provavelmente o grande desafio a ser vencido pelas *smart cities*. A escassez de recursos financeiros (na Europa principalmente, devido a recente crise financeira) combinado ao alto custo das soluções tecnológicas para *big data* agravam a situação. Outro fator apontado é a ausência de claros modelos de negócio para se implantar uma *smart city*. Para superar estes desafios os autores propõem as seguintes

soluções:

- **Desafio Político:** é necessário criar departamentos de *smart cities* nos governos (assim como existem os departamentos de TI nas organizações, por exemplo). Essa ação diminuiria a complexidade na comunicação entre os *stakeholders* das *smart cities* e os diversos setores envolvidos (e atualmente não integrados), como saúde, meio ambiente, segurança, ciência e tecnologia por exemplo. Isso facilitaria também, segundo eles, o processo de tomada de decisão, que seria centralizado neste setor apenas, e a aproximação entre governo e iniciativa privada (e. g. indústria de software, pequenos e médios desenvolvedores, produtores de dispositivos para IoT), que possuiriam seu próprio canal;
- **Desafio Tecnológico:** é preciso implantar plataformas tecnológicas que consigam integrar os dados de diversos setores (com diferentes formatos e esquemas) e que consigam lidar com o grande volume de dados produzidos (oriundos desde sensores da *IoT* até de redes sociais e *open government data* por exemplo). Para este fim, uma solução é utilizar as já citadas grandes fornecedoras de serviços e produtos, que tem desenvolvido suas próprias soluções, ou desenvolver uma solução própria. Este desafio ainda é difícil de vencer, pois o custo de implantar ou desenvolver estas tecnologias é alto. Porém, as soluções para o desafio financeiro, listadas a seguir, podem auxiliar a trazer alívio financeiro para esta tarefa;
- **Desafio Financeiro:** este desafio traz uma oportunidade interessante. Segundo os autores, é possível atrair investimentos privados e filantrópicos nesta área (pelas questões ambientais ou de qualidade urbana, por exemplo) se for usada uma estratégia inteligente. É proposto por eles então um processo em três fases para atrair investimentos externos ao governo:
 - **Fase 1:** focar em serviços que tragam retorno a curto prazo e que ofereçam informações com maior utilidade à vida dos cidadãos destas cidades. Aplicativos para segurança pública ou de trânsito, por exemplo, (como o recente Waze), que tragam melhoras rápidas e funcionalidades úteis. Se os governos executarem esta fase com sucesso, pode-se gerar mais rapidamente um fluxo de caixa para novos investimentos;
 - **Fase 2:** focar em serviços com retorno financeiro a longo

prazo. Espera-se que iniciativas privadas já sejam atraídas pelo fluxo de caixa e dados gerados na fase anterior. Aplicativos que colem e façam a mineração de dados são um exemplo interessante para esta fase;

- **Fase 3:** disponibilização dos dados e alguns serviços gerados pelas duas fases anteriores, através de uma plataforma para acesso. Fazendo isso, as *smart cities* podem gerar novas receitas, cobrando (ou não) pelo uso dos dados. Segundo os autores, o acesso a esta plataforma pode ser gerenciado pelas seguintes abordagens ou modelos de negócio:
 - **Abordagem *App-Store-Like*:** os desenvolvedores se inscrevem e acessam os dados através o uso de APIs, para que estes criem seus próprios aplicativos. Pode ser inclusive cobrada uma taxa para este acesso. Os aplicativos desenvolvidos podem ser depois disponibilizados em repositórios de comercialização de aplicativos para dispositivos móveis, como a Google Play, para Android ou outras *App Stores*, por exemplo;
 - **Abordagem *Google-Maps-Like*:** serviços que necessitem de dados com alta granularidade, confiabilidade e veracidade, exemplificando, como em aplicativos de trânsito, podem ter uma taxa cobrada conforme o nível de detalhes ou frequência de acesso aos dados. Um exemplo deste tipo de cobrança é a API *Google Prediction*, que cobra conforme a frequência de consulta aos dados aumenta;
 - **Abordagem *Open-Data*** (assim chamado pelos autores): as cidades podem fornecer acesso aos dados sem cobrança de taxas. Os autores, porém crêem que esta abordagem não é financeiramente viável e precisaria de subsídio financeiro governamental.

Estes três modelos de negócio e o processo trifásico analisado reúnem desafios e apontam direções para projetos desta natureza, além de evidenciar o papel dos desenvolvedores e indústria de software como estratégico. Logo, é necessário que este se conheça quais características que os artefatos de software devem possuir para atuar nas *smart cities*. Uma abordagem interessante surge das características existentes nos atuais dispositivos móveis combinados à IoT. Sabe-se que a internet das coisas é uma importante fonte de *big data* (CHEN; MAO;

LIU, 2014) e é uma parte essencial da infraestrutura das *smart cities*. Logo, além dos tradicionais sensores das *smart cities* (em redes de água, calculando a vazão, por exemplo), os atuais dispositivos móveis (*smartphones e tablets*, por exemplo) podem atuar também como sensores, coletando e gerando dados. Eles podem produzir informações sobre o contexto de um usuário em uma determinada situação, possibilitando a criação de uma nova geração de aplicativos baseados em *big data* (DOBRE; XHAFA, 2013). Um simples exemplo deste conceito seria uma aplicação que conseguisse coletar a localização atual de um determinado usuário, através do GPS de seu dispositivo móvel, e combinasse com suas preferências (preferências por certos tipos de produtos). Este aplicativo conseguiria enviar informações ao seu usuário sobre onde encontrar lojas de interesse mais próximos ao local em que ele estivesse em um determinado momento. Esta ideia está relacionada ao conceito de computação sensível ou perceptível ao contexto, onde os computadores podem perceber uma determinada situação e reagir a ela (ROBLES; KIM, 2010). Nesta linha, segundo Gartner (2009), perceber e reagir ao contexto de um usuário será tão influente em *apps* para dispositivos móveis quanto as máquinas de busca são para a web. Estes aplicativos representam a ideia de que cidades inteligentes são cidades que agem e reagem aos acontecimentos de seu cotidiano (NEIROTTI et al., 2014). Isso direciona e dispara uma gama de oportunidades de desenvolvimento de software (ou aplicativos) para *smart cities*. Dobre e Xhafa (2013) chamam este tipo de software como a nova geração de aplicações, e expõem quais características devem ser consideradas como requisitos por elas:

- Mobilidade e localização: as aplicações devem auxiliar o usuário a encontrar informações sobre lugares próximos a ele;
- Proximidade: a quantidade de dados em determinadas situações pode ser alta demais, e os dispositivos móveis tem capacidade de armazenamento e processamento limitada. É preciso que se filtrem os dados, trazendo apenas dados geograficamente mais próximos ao usuário;
- Garantia *Real-time*: as informações precisam ser atuais. O usuário não deve receber informações obsoletas (notícias sobre acontecimentos do mês passado, por exemplo);
- Suporte para erros de comunicação: nenhuma aplicação para *smart cities* deve assumir que o usuário está sempre conectado à Internet, ou com uma conexão que nunca caia;
- Descoberta de novas fontes: as aplicações devem conseguir

descobrir novas fontes de dados, como outros sensores e serviços externos. Isso aumentaria a possibilidade de informações e funcionalidades úteis (interação com prédios inteligentes poderiam abrir uma porta ou acionar o elevador, por exemplo);

- Escalabilidade: devem ser escaláveis e possuir histórico de armazenamento, para permitir futuras minerações. Isso permite futuras aplicações de mineração de dados muito interessantes. O problema aqui seria a baixa capacidade de armazenamento dos dispositivos, mas que poderia ser resolvido com *cloud computing*.

Dois outros trabalhos relacionados que auxiliam a compreender a questão das *smart cities* são o de Letaifa (2015) e Zygaris (2012). Zygaris (2012) propõe um framework de referência para o planejamento de *smart cities*. Este framework agrupa componentes tecnológicos e questões econômicas, ambientais e sociais a serem consideradas por estas cidades. Isto auxilia muito a identificar os componentes chaves de uma *smart city*, porém, o framework não explica as relações entre estes componentes a sua função no ambiente das *smart cities*. Letaifa (2015) reforça a importância da interdisciplinaridade nas pesquisas sobre *smart cities*. O autor afirma que as pesquisas atuais precisam avançar neste sentido, pois ainda pesquisa-se muito este tema de forma disciplinar (no sentido contrário de interdisciplinar), sem considerar a integração entre as diferentes disciplinas envolvidas, como por exemplo, a economia nestas cidades, sua relação com o desenvolvimento urbano, com questões políticas, ecológicas e educacionais, dentre outras.

2.7 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Conforme visto na revisão sobre *big data*, este complexo fenômeno corresponde ao aumento mundial de volume de dados, que estão em diversos formatos, estruturados ou não, modelados conforme diversos domínios diferentes, e este volume cresce velozmente. Os autores citados na revisão sobre *big data* e sobre *smart cities* apontam que este fenômeno pode contribuir para iniciativas de transformar uma cidade em uma *smart city*, além de ser também uma consequência deste processo. Dentro deste contexto, as abordagens estudadas para *smart cities* baseadas em *big data* enfrentam diversos desafios atualmente, tanto administrativos quanto tecnológicos. Do ponto de vista tecnológico, é verificado que as principais características do fenômeno *big data* tornam difícil a sua análise e extração de conhecimento. Em

outras palavras, para analistas e gestores destas cidades, é complexa a atividade de organizar todo este montante de dados, tratar todas suas diversas características (formatos, esquemas, periodicidade de produção, atualização em tempo real) e transformá-los em informação, para que sirvam de insumo para a geração de conhecimento para estas cidades, além de se preocuparem com a confiabilidade das fontes de dados. Logo, para auxiliar esta tarefa, propô-se um modelo conceitual que foca em *big data* e *open data* e que é orientado ao problema das *smart cities*, conforme seus domínios de atuação, para que seja utilizado pelos gestores destas cidades, e demais analistas ou pesquisadores interessados no tema.

Sobre as definições de *smart cities*, é importante ressaltar que se percebe a alta influência das TICs na medida em que elas devem ser aplicadas para melhor apoiar os serviços (exemplos: mobilidade urbana e segurança pública) e patrimônios (capital humano e social) essenciais em uma cidade. Além disso, fatores humanos, sociais e políticos, como a educação, cultura e inclusões social e digital por exemplo, são considerados chaves para um desenvolvimento eficiente nas atuais cidades e na resolução de seus problemas. Como uma das justificativas para se investir em capital humano, está o fato de que as *smart cities* enfatizam o conhecimento como fator estratégico para gerar inovação e competitividade nas suas economias, contribuindo para seu desenvolvimento e prosperidade, tornando-as mais eficientes e eficazes (*smart*) em seus processos e áreas chaves. Contribuindo com o fenômeno *big data*, e indo de encontro à abordagem das *smart cities*, está o movimento de dados abertos (*open data*), mais precisamente, dados abertos governamentais. Tal movimento apresenta a geração de conhecimento como uma de suas justificativas para sua adoção pelas cidades, uma vez que, com mais *datasets* (conjuntos de dados) abertos e disponíveis para utilização, maior é a possibilidade de gerar conhecimento para estas cidades. Dados (abertos e governamentais) sobre o funcionamento das cidades podem chamar a atenção de especialistas de diversos domínios para que estes possam gerar seus *insights* sobre problemas complexos que as cidades enfrentam (ex.: mobilidade urbana, saúde). Em outras palavras, é possível gerar mais conhecimento através da reprodução destes dados em outros setores da sociedade, como instituições de ensino e pesquisa, indústria, empresas e na sociedade em geral. Como outra consequência e benefício desta abordagem destaca-se o aumento da transparência, um maior controle social e um maior engajamento político da sociedade.

Vilajosana et al. (2013) criticam a abordagem *open data*, porém é

necessário discordar. O movimento *Open Data* têm a capacidade de potencializar o uso do fenômeno *big data* como fonte de dados, pois quando aplicada a licença de dados abertos às fontes *big data*, seus dados passam a ter legitimidade legal para sua obtenção e uso (MANYKA et al., 2013), além de prover uma maior interoperabilidade a estes dados (ex.: RDF ao invés de PDF).

Em relação a modelos conceituais, Monteiro (2006), citando Dahlberg (1978), afirma que a formação de conceitos se desenvolve na medida em que estes são definidos, e suas composições e relacionamentos são estabelecidos. Além disso, Monteiro (2006) citando Campos (2004), o autor também afirma que " as relações entre objetos de um dado contexto formam a estrutura conceitual deste contexto...". Logo, percebe-se nestas afirmações a importância de três elementos: conceitos, relações e seu ambiente. Nesta linha, dentro de uma visão sistêmica, Bunge (2003) auxilia a modelar conceitos suas relações e sua interação com seu meio. Para Bunge (2003) qualquer sistema pode ser modelado com a quádrupla CESM – *Component* (Componente), *Environment* (Ambiente), *Structure* (Estrutura) e *Mechanism* (Mecanismos). Esta quádrupla, além de auxiliar a identificar os conceitos chaves a serem definidos em um determinado ambiente (ou sistema), também facilita a identificação dos papéis destes conceitos dentro do sistema estudado, bem como a entender o funcionamento deste sistema. Ou seja, com este método, é possível conceitualizar um determinado ambiente, identificando além de seus principais elementos, suas relações e interações com o meio e seus principais processos, o que vem a ser de grande utilidade para entender qualquer sistema, inclusive as *smart cities* e sua relação com o fenômeno *big data*, *open data*, seus componentes, processos e atores principais.

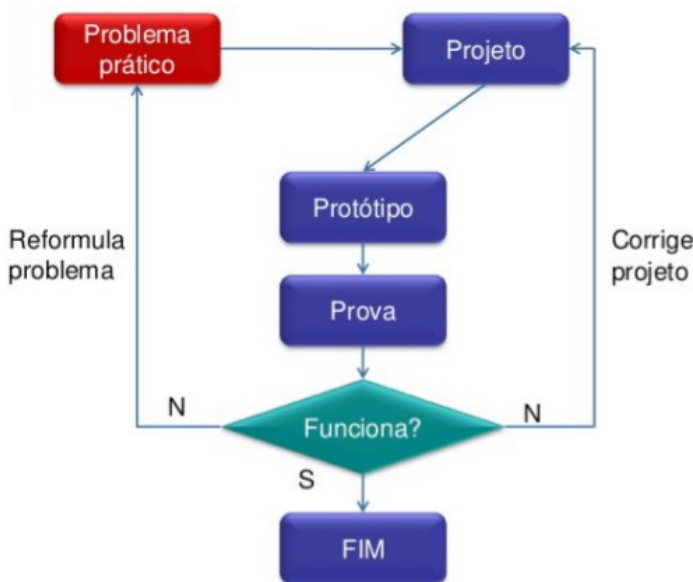
3 PROCEDIMENTOS METODOLÓGICOS

Neste capítulo são apresentadas as características desta pesquisa, sua classificação conforme seu objetivo e seus procedimentos, bem como os procedimentos executados e resultados obtidos.

3.1 DELINEAMENTO DA PESQUISA

Conforme afirmado anteriormente, esta pesquisa caracteriza-se como uma pesquisa tecnológica, pois está preocupada em produzir um artefato funcional como resultado final (CUPANI, 2006; FREITAS JUNIOR, et al. 2012), conforme ilustrado na Figura 6.

Figura 6: Ciclo mínimo de produção tecnológica



Fonte:Cupani, citando Bunge (1985).

Quanto aos procedimentos metodológicos adotados, esta pesquisa classifica-se como exploratória/bibliográfica (GIL, 2002). Seu delineamento busca implementar esta classificação, conforme pode ser visto no Quadro 4.

Quadro 4: Delineamento da pesquisa

Procedimentos Metodológicos
Passo 1: Pesquisa bibliográfica: leitura analítica e leitura interpretativa
Passo 2: Identificação dos atores (<i>stakeholders</i>) e seus papéis
Passo 3: Definição de Requisitos
Passo 4: Desenho e descrição do Modelo
Passo 5: Desenvolvimento dos artefatos
Passo 6: Aplicação do Modelo em um cenário de uso

Fonte: o autor.

Abaixo seguem os resultados do Passo 1 e em seguida os demais passos.

3.2 PESQUISA BIBLIOGRÁFICA

Segundo Botelho (2012), a revisão da literatura é o primeiro passo para a construção de conhecimento, pois é por meio desse processo que são identificadas lacunas e oportunidades de pesquisas. Na pesquisa bibliográfica executada, primeiro passo desta pesquisa, foram utilizados as palavras chaves (termos de busca) *big data*, *smart cities*, *open data* e modelos conceituais (*conceptual models*), em português e inglês, isoladamente e combinadas entre si, como é demonstrado no Quadro 5. Foi utilizado como base as coleções virtuais do Portal Capes, Biblioteca Universitária da UFSC, livros, relatórios e *white papers* de grandes empresas do setor (com destaque para os relatórios produzidos pela IBM e McKinsey). No caso de periódicos, foram filtrados apenas os revisados por pares e publicados após 2000 (até 2014). Foi buscado também na interface *web IEEE Xplore*, que publica materiais de computação dentre outras áreas, porém não oferece o filtro de revisão por pares. Os resultados obtidos podem ser visualizados no Quadro 5.

Quadro 5: Consultas em bases bibliográficas

Termo de busca	Base de dados						
	Scopus (Elsevier)	One File - GALE	Web of Science	ProQuest	Computer Inform	IEEE	Outras

					mation Syste ms Abstra cts		
<i>Big Data</i>	1940	1587	1782	1411	494	298	3501
<i>Smart Cities</i>	180	126	152	90	41	38	317
<i>Open Data</i>	540	301	429	312	116	44	1105
<i>Big Data + Smart Cities</i>	23	25	9	18	4	3	47
<i>Big Data + Open Data</i>	45	34	22	44	13	6	82
<i>Open Data + Smart Cities</i>	7	5	2	6	3	----	17
<i>Big Data + Open Data + Smart Cities</i>	1	1	1	1	----	----	----
<i>Conceptual Models</i>	2612	1266	2721	426	185	430	5566
<i>Conceptual Model + Smart Cities</i>	1	----	1	----	----	2	1
<i>Conceptual Model + Big Data</i>	4	----	----	2	----	14	3
<i>Conceptual Model + Open Data</i>	1	----	----	----	----	----	----
<i>Conceptual Model + Big Data +</i>	----	----	----	----	----	----	----

<i>Open Data</i>							
------------------	--	--	--	--	--	--	--

Fonte: o autor.

A seguir é apresentada a proposta desta dissertação.

4 A PROPOSTA

No Capítulo 4 é apresentada e detalhada a proposta desta dissertação. São descritos os requisitos levantados e o modelo conceitual desenvolvido. Foi elaborado um desenho do modelo e uma descrição textual de seu funcionamento, bem como detalhado seus componentes e camadas, conforme segue.

4.1 REQUISITOS LEVANTADOS

Para auxiliar a desenvolver o Modelo Conceitual proposto foram utilizadas como ferramentas o processo de desenvolvimento de ontologias OntoKEM (RAUTENBERG; TODESCO; GAUTHIER, 2009), o framework de Zachman (ZACHMAN, 2003), e o modelo CESM para construção de sistemas teleológicos (BUNGE, 2003). Como base teórica para ser usada por estas ferramentas foram utilizados os diversos autores e trabalhos relevantes da área, sendo os principais apresentados no capítulo de fundamentação teórica desta dissertação.

O OntoKEM é um processo de desenvolvimento de ontologias que combina algumas práticas de metodologias para construção de ontologias, como a *Ontology Development 101* (processo iterativo para desenvolvimento de ontologias), *On-to-Knowledge* (definição do domínio e o escopo da ontologia; utilização de questões de competência), e *Methontology* (rico conjunto de artefatos de documentação) (RAUTENBERG et al., 2010). Este processo prevê quatro fases de desenvolvimento: especificação (identificar propósito, identificar escopo, considerar reuso, identificar fontes), aquisição de conhecimento (gerar questões de competência, listar termos, definir termos, definir propriedades, agregar elementos, definir relações, definir restrições, criar instâncias), implementação (valorar propriedades, valorar relações, valorar restrições) e avaliação (avaliar segundo fontes, avaliar segundo requisitos, avaliar com usuários) (RAUTENBERG et al., 2010). A escolha pelo OntoKEM se justifica pelo fato de ontologias serem comumente usadas para representar o conhecimento e suas estruturas (RAUTENBERG; TODESCO; GAUTHIER, 2009), e um dos objetivos do Modelo proposto é possibilitar a geração de conhecimento para as *smart cities*. Dentre as práticas previstas pelo OntoKEM, foi focado principalmente nas práticas de “Questões de Competência”, “Listar Termos”, “Definir Termos”, “Agregar Elementos” e “Definir Relações”. As Questões de Competência representam os requisitos para

desenvolvimento do Modelo. São questões às quais o Modelo deve auxiliar a responder. Elas seguem listadas a seguir e são respondidas no capítulo 4.3 (descrição do funcionamento do modelo). Percebe-se que elas derivam principalmente das questões de pesquisa proposta. As questões foram agrupadas em três camadas, conforme segue:

- Camada Conceitual (*Conceptual Layer*):
 - O que é uma *smart city* (*What is a smart city*)?
 - Quais são os termos associados (*What are the related terms*)?
 - Como estes termos diferenciam-se entre si (*How are the many terms different from each other*)?
- Camada de Domínios e Indicadores (*Domain Layer*):
 - Quais dados coletar (*What data to collect*)?
 - Qual o escopo a ser abordado (*What is the scope or subjects to analyse*)?
 - Como extrair informações sobre o mecanismo (os seis domínios) das *smart cities*, para analisar seu funcionamento (*How to extract information about the mechanism/ six domains of smart cities, to analyze its operation*)?
 - Quem são os principais atores nas *smart cities* (*Who are the stakeholders*)? Como eles influenciam as cidades inteligentes (*How do they influence the smart cities*)?;
- Camada de Dados (*Data Layer*):
 - Quais são as principais fontes de dados (*What are the main data sources*)?
 - Dados sobre quais domínios pode-se encontrar nestas fontes (*What smart cities domains data can be found*)?
 - Em quais formatos são encontrados estes dados (*What are the main content format of the data sources*)?
 - Quais suas principais características (*What are its main characteristics*)?
 - Qual a periodicidade em que eles são produzidos (*What is the data production velocity*)?

Para auxiliar o desenvolvimento das Questões de Competência listadas acima foi utilizado também o framework de Zachman para arquiteturas corporativas, que de acordo com Zachman (2003), é um modelo que realiza seis palavras interrogativas primitivas (por que, como, o que, quem, onde e quando) às quais, se respondidas, auxiliam a

entender ideias complexas e objetos de estudo de diversos domínios.

Após elaborar os requisitos do modelo, em forma de Questões de Competência, foi necessário respondê-las. Para isso, foi utilizado modelo CESM (BUNGE, 2003), aplicando-o à fundamentação teórica desta dissertação. Logo, foi pensado nas *smart cities* como um sistema, e então, foram identificados seus itens e fluxos de entrada e saídas, com auxílio do método CESM. Bunge (2003) afirma que qualquer sistema pode ser modelado conforme a quádrupla de elementos “*Composition*” (composição), “*Environment*” (ambiente), “*Structure*”, (estrutura) e “*Mechanism*” (mecanismo). Esta caracterização corresponde ao modelo CESM (BUNGE, 2003), que pode ser utilizado como método para modelar sistemas teleológicos: sistemas projetados em uma abordagem *Top-down*, guiados por um objetivo final (SCHMITT; DOROW; SOUZA, 2012 apud BOSCO, 2006), nos quais se incluem, dentre vários exemplos, sistemas em geral, como empresas, sistemas de transporte, sistemas de informação, dentre outros. Dentre os elementos e componentes modelados, foram identificados também os principais atores das *smart cities*, seus papéis neste ambiente, as relações entre si, bem como o principal mecanismo destas cidades. Os resultados seguem descritos na seguir.

4.2 OS PAPÉIS DEFINIDOS E SUAS FUNÇÕES

Conforme mencionado anteriormente, Bunge (2003) afirma que qualquer sistema pode ser modelado conforme a seguinte quádrupla de elementos: “*Composition*” (composição), “*Environment*” (ambiente), “*Structure*”, (estrutura) e “*Mechanism*” (mecanismo). Esta caracterização corresponde ao modelo CESM (BUNGE, 2003), que pode ser utilizado como método para modelar sistemas teleológicos: sistemas projetados em uma abordagem *Top-down*, guiados por um objetivo final, nos quais se incluem vários tipos de sistemas, como empresas, sistemas de transporte, sistemas de informação dentre outros exemplos (SCHMITT; DOROW; SOUZA, 2012 apud BOSCO, 2006).

Para melhor compreender este conceito, considerando o exemplo do Sistema Solar, a quádrupla CESM pode ser representada da seguinte maneira, e conforme ilustrado no Quadro 6:

- Composição (*Composition*): é formado pela coleção de componentes do sistema. Exemplo: sol, planetas e asteroides;
- Ambiente (*Environment*): coleção de itens que afetam ou são afetados pelos componentes ou o sistema, mas não fazem parte do sistema. Exemplo: demais corpos celestes;

- Estrutura (*Structure*): coleção de relações entre os componentes (endoestrutura) ou entre esses e seu ambiente (exoestrutura). Exemplo: Forças gravitacionais;
- Mecanismo (*Mechanism*): coleção de processos do sistema que o fazem se comportar da maneira que é. Exemplo: Os Processos de emissão e absorção de luz, combinação.

Quadro 6: Ilustração através de exemplos de aplicação do modelo CESM de Bunge

Sistema	Composição	Ambiente	Estrutura	Mecanismo
Sistema Solar	Sol, planetas e asteroides	Demais corpos celestes	Forças gravitacionais	Os Processos de emissão e absorção de luz, combinação.
Empresa	Funcionários e Gerentes	Mercado e Governo	Relações de trabalho entre membros da empresa e entre esses e o ambiente	Atividades que resultam nos produtos da empresa

Fonte: Adaptado de Kern (2011).

Em seguida, foi feita a identificação dos elementos (composição, ambiente, estrutura e mecanismo) de *smart cities* baseadas em *big data* e *open data*, mapeados com base nos resultados encontrados na literatura e agrupados no modelo conceitual já apresentado. Abaixo, seguem listados o resultado deste mapeamento. Os termos foram mantidos em inglês, seu idioma original, mas traduzidos ao seu lado.

- **Composição:** os principais atores, com papéis chaves no processo principal das *smart cities* (que é elevar a qualidade de vida de seus habitantes):
 - 1- *City administrators*: o governo ou os administradores da cidade;
 - 2- *Civil Society*: cidadãos ou sociedade civil;
 - 3- *Industry*: indústria em geral (*industry*), e três subclasses de atores de destaque desta categoria: indústria de TI (*IT industry*), fornecedores de tecnologia para *IoT* (*IoT Providers*) e desenvolvedores (*developers*);

- 4- *University*: instituições de ensino e pesquisa, como universidades, escolas, faculdades e institutos de pesquisa por exemplo;
- 5 e 6 -*Utilities* e *Policies*: serviços públicos inteligentes (*utilites* ou *smart utilities*) e políticas públicas (*policies*), respectivamente. Devido a forte importância dada por diversos autores ao papel destes elementos chaves, estes componentes seguem destacados como atores (DOBRE; XHAFA, 2013; VILAJOSANA et al., 2013). Estes dois atores são criados pelo ator *City Administrator* (administradores de cidades) e se relacionam com os demais atores deste sistema, produzindo dados e criando ações, respectivamente. Exemplo: dados produzidos pelos sensores de uma rede de tratamento de esgoto de uma cidade (*utilities*, domínio *hard*) e ações para melhorar a consciência ambiental da população nestas cidades, ou para dispor os seus dados governamentais em formato aberto (*policies*, domínio *soft*);
- 7 e 8: *Data* e *Services*: os dados e os serviços também são considerados componentes chaves deste sistema. De maneira geral, os dados são produzidos pelas *smart utilities*, mas também são produzidos por atores como *City Administrators* (exemplo: dados governamentais), *Civil Society* (exemplo: postagens em sites, redes sociais, *living labs*), servindo como insumo básico para a promoção de conhecimento (promovida principalmente pelo ator *University*), e para a criação de novos serviços e consequentemente, gerando inovação e desenvolvimento econômico (realizados principalmente pelos atores *Industry* e *Developers*). O componente *Services* geralmente é realizado em uma parceria PPP (Parceria Público-Privada), onde a iniciativa privada utiliza os dados públicos para gerar inovação e desenvolvimento econômico (VILAJOSANA et al., 2013). Para isso, conforme exposta na atual proposta, deve-se utilizar dados governamentais em formato aberto, podendo estes pertencer ao fenômeno *big data* e outras fontes (ex: *linked open data*, dados de sensores). O detalhamento das diversas fontes e suas características (ex.: estruturado, semiestruturado...) são detalhados na Camada de Dados desta proposta;

- **Ambiente:** a Cidade em Questão, Governos Federal, Estadual, demais Governos e instituições (outras entidades relacionadas que possam ser parceiros comerciais, via intercâmbio cultural ou outras formas de relacionamento por exemplo). Governos Municipais de cidades limítrofes também entram neste grupo, pois podem interagir com a Cidade em Questão (exemplo: um sistema de trânsito intermunicipal);
- **Mecanismo:** são as principais ações que levam as *smart cities* a realmente se tornarem *smart*, auxiliando-as em seu objetivo principal, que é elevar a qualidade de vida de seus habitantes. Estas ações são agrupadas em seis domínios (*domains*), ou características genéricas de Giffinger et al. (2007), que definem uma *smart city*. As ações destes domínios devem ser implementados pelos atores das *smart cities*. São estes os seis domínios: *Smart Environment* (Meio Ambiente), *Smart Mobility* (Mobilidade), *Smart People* (Capital Humano e Social), *Smart Economy* (Economia), *Smart Government* (Governo Transparente) e *Smart Living* (Qualidade de Vida).
- **Estrutura:** neste grupo, foram identificadas as relações entre os atores e seu ambiente (exosestrutura) e dos componentes entre si (endoestrutura). Exemplo: os administradores da cidade (*City Administrators*) equipam com sensores os seus serviços públicos (*Utilities*), que são fornecidos pela indústria de *IoT* (*IoT Industry*). Os cidadãos (*Civil Society*), que além de usarem estes serviços, produzem dados, postando em redes sociais, como sua opinião da sobre uma determinada ação política por exemplo. As instituições de ensino e pesquisa auxiliam a promover o conhecimento baseado nestes dados, e a indústria (*Industry*), juntamente com os desenvolvedores, geram o desenvolvimento econômico, através da inovação de novos produtos e serviços (*Services*), baseados nestes dados (*Data*). Isso acaba formando um ciclo virtuoso de geração de dados, informação e conhecimento, que pode ser visto na Figura 13. O componente *Data* serve de base (***is base***) para o componente *Services* (pois os dados são baseados nos serviços, que entregam informações aos usuários). Este por sua vez (*Service*), transforma os dados em informações e geram inovações (***Innovate***). Estas duas relações em negrito (***Is Base*** e ***Innovate***) seguem listadas no Quadro 7, assim como as demais estruturas mapeadas. Abaixo também **em negrito**, estas relações são

exemplificadas e relacionadas a seus atores (componentes), para facilitar o seu entendimento:

- *City Administrators* (Administradores das Cidades): estes atores se relacionam com o meio na medida em que criam as políticas públicas (**Create Policies**, exemplo: políticas de proteção ambiental), regulam o uso dos dados (**Regulate Data**, exemplo: publicação de dados governamentais em formato aberto), e aperfeiçoam os serviços públicos (**Improve Utilities**, exemplo: equipando-os com sensores e demais controles de eficiência com o uso de TICs);
- *Civil Society* (sociedade civil): suas principais relações com os demais componentes e ambiente são: produção de dados (**Produce Data**), utilização das *Utilities* (**Use Utilities**, utilizam os serviços públicos inteligentes, exemplo: redes elétricas inteligentes) e consumo de *Services* (**Use Services**, os serviços desenvolvidos pela *IT Industry* e *Developers*, exemplo: aplicativo Waze). Para melhorar o entendimento deste item, segue este exemplo: imagine cidadãos interagindo em aplicativos de trânsito, ou fazendo postagens em redes sociais *online* (ex.: Twitter), produzindo informações sobre a cidade, emitindo sua opinião sobre as políticas públicas de proteção ambiental, ou informando avisos em aplicativos de denúncia pública sobre falhas em redes de abastecimento de água. Estes mesmos cidadãos podem também utilizar serviços proprietários (*Services*), como o citado Waze, que utiliza dados geográficos e sobre o trânsito das cidades;
- *Industry*: de maneira geral, sua relação com o meio é gerar o desenvolvimento econômico das cidades (**Generate Economic Growth**). Além da indústria em geral, destacam-se os: 1- os fornecedores de tecnologia para *IoT* (*IoT Providers*), que fornecem as tecnologias necessárias para as *smart utilities* (**Provide IoT Infrastructure**), exemplo: sensores; 2- a Indústria de TI (*IT Industry*) e Desenvolvedores (*Developers*) em geral fornecem demais soluções de software e serviços para estas cidades (**Produce Services**), gerando inovação, como por exemplo, aplicativos para melhorar a mobilidade urbana, também auxiliando o crescimento econômico;
- *University*: sua principal relação com o meio e os demais

componentes é promover o conhecimento nestas cidades (**Promote Knowledge**), elevando seu capital humano e social. Além disso, podem também relacionar-se com estes componentes ao fornecer conhecimento sobre as cidades, através de seus especialistas e pesquisadores. Ou seja, ao analisar os dados das cidades, pesquisadores e/ou especialistas de diversos domínios podem auxiliar as cidades em seus problemas urbanos, e assim inferir e gerar conhecimento sobre as cidades. Exemplo: um geógrafo analisando um fenômeno natural ou um sociólogo estudando índices de crime, ao usar dados públicos governamentais e abertos). Isso promoveria também o compartilhamento de conhecimento;

- *Policies*: este componente implementa as ações (**Implement Actions**) chaves criadas principalmente pelo componente *City Administrator*, mas também podem ser iniciadas por movimentos sociais, ONGs e até pelo ator *Civil Society*, através do engajamento político dos cidadãos destas cidades (exemplo: um cidadão ou uma ONG propondo projetos de lei para proteção e criação de reservas ambientais). Suas ações estão mais relacionadas com o ambiente, agindo principalmente (mas não exclusivamente) em domínios considerados *soft* (*smart people, smart economy, smart government*). Exemplos: a criação de transparência governamental através da abertura de dados, ações de inclusão social e digital, investimentos em educação, proteção ambiental, projetos de lei criados pela sociedade civil, dentre outros;
- *Utilities* (Serviços públicos inteligentes ou *smart utilities*): Produzem dados sobre o funcionamento da Cidade em Questão. A sua principal interação com os demais atores está na relação **“Produce (City) Data”**, pois os componentes *Civil Society*, *City Administrator* e *Industry* consomem esses dados, para produzir ou utilizar *Services*.

Para melhor compreender a aplicação do modelo CESM às *smart cities* é necessário tratá-las como um sistema, cujo objetivo principal é elevar a qualidade de vida de seus habitantes. Este sistema possui ações agrupadas em seis principais domínios de atuação (*Smart Environment, Smart Mobility, Smart People, Smart Economy, Smart Government e Smart Living*), representando seus principais processos (*Mechanism*),

cujos resultados levam ao seu objetivo final, que é promover a qualidade de vida neste sistema. Além disso, seus atores são os principais componentes do sistema (*Components*), que atuam diretamente nestes processos. As inter-relações entre estes componentes formam a estrutura deste sistema (*Structure*: 1.1-*Create Policies* – Criam Políticas Públicas; 1.2- *Regulate Data* – Regulam o uso dos dados; 1.3 - *Improve Utilities* – Melhoram os Serviços Públicos; 2.1 -*Produce Data* – Produzem Dados; 2. 2 - *Use Utilities* – Utilizam os Serviços Públicos; 2.3- *Use Services* – Utilizam os Serviços PPP – Parceria Público Privado; 3 - *Generate Economic Growth* – Desenvolvem a Economia; 3.1 e 3.2 - *Produce Services* – Desenvolvem Serviços PPP e Inovação; 3.3 - *Provide IoT Infrastrucutre* – Fornecem Infraestrutura de IoT; 4 - *Promote Knowledge* – Promovem o Conhecimento; 5 - *Produce (City) Data* – Produzem Dados sobre a cidade; 6 - *Implement Policies* – Implementam as Políticas Públicas;7 e 8: *Is Base* e *Innovate* – São Base e Inovam), conforme Quadro 7 e na Figura 7.

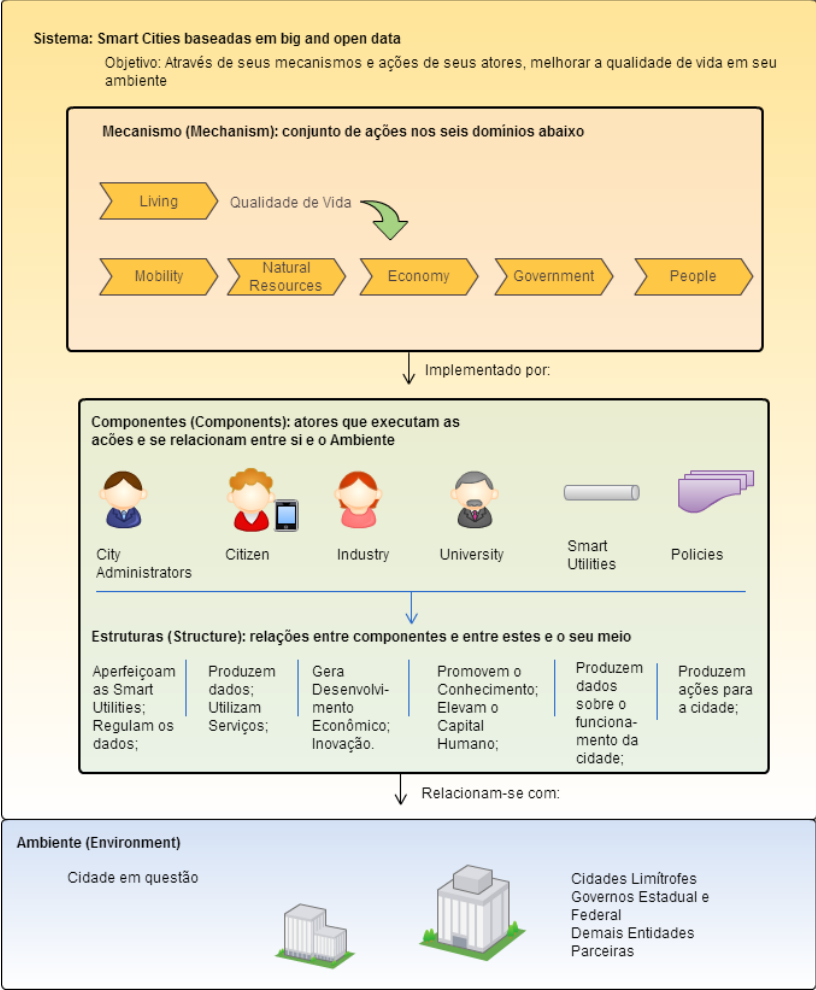
Quadro 7: Aplicação do modelo CESM de Bunge para desenvolver o Modelo

Sistema	Composição	Ambiente	Estrutura	Mecanismo
Sistema modelado com o modelo CESM:	Definição: Coleção de componentes do sistema, conforme segue abaixo:	Definição: Coleção de itens que afetam ou são afetados pelos componentes ou o sistema, mas não fazem parte do sistema:	Definição: Coleção de relações entre os componentes (endoestrutura) ou entre esses e seu ambiente (exoestrutura):	Definição: Coleção de processos do sistema que o fazem se comportar da maneira que é:
<i>Smart Cities</i> , com foco <i>big data</i> e <i>open data</i>	Os principais atores e outros componentes chaves: 1-City Administrator,	1-A Cidade em Questão; 2-Governos Estaduais, Federal, cidades limítrofes, e demais Entidades	Relações entre os principais atores entre si e o seu meio: 1.1- <i>Create Policies</i> ; 1.2- <i>Regulate</i>	Processos dos seis domínios chaves: <i>Smart Environment</i> (Meio Ambiente), <i>Smart Mobility</i> (Mobilidade), <i>Smart People</i>

Sistema	Composição	Ambiente	Estrutura	Mecanismo
	2-Civil Society;; 3- Industry; 3.1 -IT industry, 3.2- Developers; 3.3 – IoT Providers; 4-University; 5-Utilities; 6-Policies; 7 e 8: Data e Services.	parceiras (ex.: cidades estrangeiras com parceiras econômicas ou de intercâmbio).	Data; 1.3 - Improve Utilities; 2.1 -Produce Data; 2. 2 - Use Utilities; 2.3- Use Services; 3 - Generate Economic Growth; 3.1 e 3.2 - Produce Services; 3.3 - Provide IoT Infrastructure; 4 - Promote Knowledge; 5 - Produce (City) Data; 6 - Implement Policies 7 e 8: Is Base e Innovate.	(Capital Humano e Social), Smart Economy (Economia), Smart Government (Governo) e Smart Living (Qualidade de Vida).

Fonte: o autor

Figura 7: Aplicação do Modelo CESM às *Smart Cities* Baseadas em *Big Data* e *Open Data*



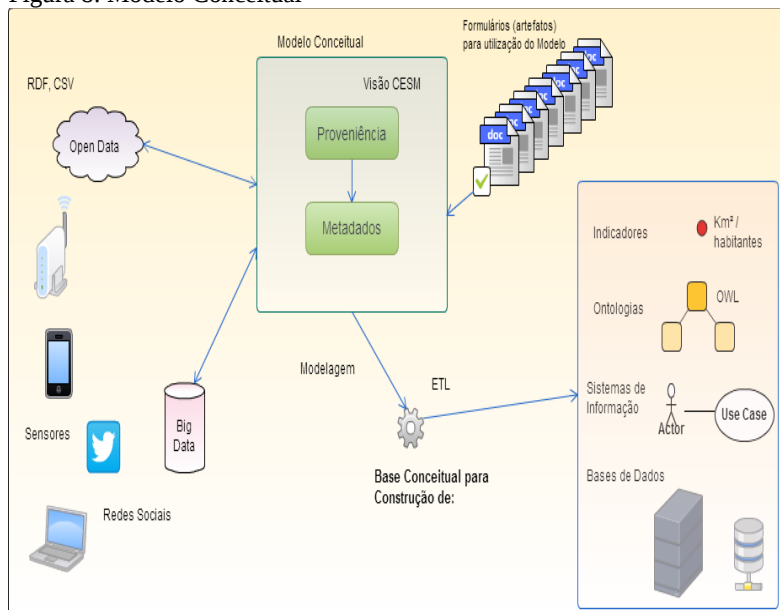
Fonte: o autor

Em seguida, é apresentado o desenho que visa explicar o Modelo proposto, e seu funcionamento é descrito na sequência.

4.3 DESENHO DO MODELO E DESCRIÇÃO DE SEU FUNCIONAMENTO

Após levantar os requisitos do Modelo, em forma de Questões de Competência, aplicar o modelo CESM para identificar os principais conceitos, atores, e suas relações no sistema *smart cities*, foi desenvolvido o desenho que explica o Modelo proposto, ilustrado na Figura 8.

Figura 8: Modelo Conceitual



Fonte: o autor.

O Modelo proposto implementa duas tarefas: apontar fontes de dados e catalogar a proveniência dos dados oriundos destas fontes (*big data* e *open data*); ou a identificar a necessidade de criação de bases de dados para *smart cities*, sugerindo conjuntos de metadados para cada domínio das *smart cities*, para que sejam produzidos dados e consequentemente informações baseadas nestes. Nos dois casos, os dados, existentes ou a serem produzidos, são catalogados conforme os conceitos identificados com o modelo CESM, descrito anteriormente. Esta catalogação é uma descrição de sua proveniência e de suas principais características identificadas na literatura. Esta etapa é realizada com a aplicação de formulários (artefatos produzidos e descritos no Capítulo 4.4 desta dissertação), que identificam a proveniência dos dados *big data* e *open data*. Além disso, estes artefatos

também sugerem conjuntos de metadados, agrupados pelos domínios das *smart cities*, a serem utilizados para construir bases de dados, que servirão de insumo para a construção de sistemas de informação, indicadores de desempenho e outras aplicações possíveis. Estes conjuntos foram inspirados em dados e indicadores utilizados em projetos de *smart cities* e ações implementadas nestas iniciativas. Após estas duas tarefas, os dados produzidos devem passar por processos de ETL (Extração Tratamento e Carga) para serem utilizados (e reutilizados) em sistemas de informação ou construção de indicadores. Os metadados sugeridos também podem ser usados para guiar construção de ontologias.

Para realizar as tarefas de identificação da proveniência dos dados e sugerir os conjuntos de metadados, foram agrupados em três camadas os conceitos, relações, estrutura e mecanismos, que haviam sido identificados através do modelo CESM, e que compõem o Modelo conceitual proposto. Estas camadas seguem descritas abaixo e descrevem as respostas das Questões de Competência elaboradas anteriormente.

4.3.1 Camada Conceitual

Primeiramente, focou-se em definir o conceito inicial (*smart cities*) a ser explorado pelo Modelo, e a sua diferenciação entre os termos mais comumente associados. Isso foi feito para esclarecer o conceito e evitar equivocadas interpretações do mesmo. Conforme visto na fundamentação teórica, essa tarefa se justifica pelo fato de o termo *smart cities* ter diversas definições conforme o seu contexto mas nenhuma definição formal única e de comum uso e aceitação.

Questões de competência elaboradas para construir a camada conceitual:

- O que é uma *smart city* (*What is a smart city*)?
- Quais são os termos associados (*What are the related terms*)?
- Como estes termos diferenciam-se entre si (*How are the many terms different from each other*)?

A seguir, são expostos os apontamentos da literatura para cada questão de competência da camada conceitual.

Questões de Competência: *What is a smart city* (o que é uma cidade inteligente) ou, em uma outra forma, qual a definição de *smart city* adotada?

Conforme visto, ainda não há uma definição formal que represente todas as dimensões do amplo conceito de *smart cities*. Para

preencher esta lacuna de pesquisa, utilizou-se mais de um autor para definir o conceito, sendo os principais listados a seguir.

Apontamentos/respostas escolhidas, encontradas na literatura:

- Uma cidade com bom desempenho em seis principais eixos: economia (competitividade e inovação), pessoas (capital humano e social), governança, mobilidade, meio ambiente e qualidade de vida [...]. Este conceito (cidade inteligente) geralmente se refere à busca e identificação de soluções inteligentes que permitam que as cidades modernas possam melhorar a qualidade dos serviços prestados aos cidadãos (GIFFINGER et al., 2007, traduzido e adaptado).
- *Smart city as a high-tech intensive and advanced city that connects people, information and city elements using new technologies in order to create a sustainable, greener city, competitive and innovative commerce, and an increased life quality.*”(KOMINOS, 2011).
- Cidades inteligentes incutem informações em sua infra-estrutura física para melhorar a mobilidade urbana, conservar energia, melhorar a qualidade do ar e da água, identificar problemas e resolvê-los rapidamente, recuperar-se rapidamente de desastres, coletar dados para tomar melhores decisões, utilizar seus recursos de forma eficaz em geral e compartilhar dados para permitir a colaboração entre diferentes entidades e domínios (NAM; PARDO, 2011, traduzido e adaptado).
- *Smart cities are the result of knowledge-intensive and creative strategies aiming at enhancing the socio-economic, ecological, logistic and competitive performance of cities. Such smart cities are based on a promising mix of human capital (e.g. skilled labor force), infrastructural capital (e.g. high-tech communication facilities), social capital (e.g. intense and open network linkages) and entrepreneurial capital (e.g. creative and risk-taking business activities.* (KOURTIT; NIJKAMP, 2012).

Após a definição do conceito principal foram definidos os termos mais comumente relacionados, para melhor distingui-los e tornar mais

claro o objeto de estudo. Foi identificado que o conceito central exposto possui diversas variações agrupadas por três dimensões, conforme segue abaixo, e ilustrado na Figura 9.

Questões de Competência: quais são os temas associados (*What are the related terms*)? Como eles se diferenciam entre si (*How are the many terms different from each other*)?

Apontamentos encontrados/respostas: em geral, os termos associados às *smart cities* são relativos a três dimensões destas cidades: humana, tecnológica e institucional (NAM; PARDO, 2011), conforme Figura 9. Estas dimensões definem conceitualmente os termos relacionados às *smart cities*, que, muitas vezes, são apenas variações do mesmo ou são conceitos que focam em um único aspecto destas cidades, conforme segue:

Dimensão Humana (*Human Dimension*):

- Cidades Criativas (*Creative City*): uma visão das *smart cities* que foca na criatividade como uma força crucial para o desenvolvimento de uma cidade, através das pessoas, da educação e do conhecimento como força de trabalho (NAM; PARDO, 2011 apud FLORIDA, 2002);
- Cidades Humanas (*Human City* ou *Human Smart Cities*): conceito ligado ao de cidades criativas, pelo fato de o perfil dos seus cidadãos ser mais associado ao trabalho intensivo em conhecimento e à indústria criativa (NAM; PARDO, 2011 apud GLAESER; BERRY, 2006), mas também pelo fato de centralizar suas ações nas pessoas e na sua qualidade de vida (HSU; CHAN, [2014]), entendendo e resolvendo as necessidades comunitárias através de técnicas sociais e tecnologias da informação (HUMAN SMART CITIES, 2015).
- Cidades de Aprendizado (*Learning City*): *smart cities* também são *learning cities*, na medida em que estas cidades aumentam a sua competitividade econômica na indústria o conhecimento (NAM; PARDO, 2011 apud PLUMB; MCGRAY, 2007), através de investimentos no capital humano. Neste conceito, as cidades devem fornecer aos seus cidadãos de forma inclusiva o acesso à educação (de nível básico a superior), revitalizar a educação nas famílias e comunidades, utilizar tecnologias para facilitar a educação, aumentar a qualidade e excelência na educação oferecida e impor uma cultura de aprendizado constante em seus habitantes (UNESCO, [2014]);
- Cidades do Conhecimento (*Knowledge Cities*): conceito

análogo ao termo *learning cities* (NAM; PARDO, 2011), referindo-se que as cidades são intencionalmente projetadas para utilizar o conhecimento como ferramenta que gere o bem estar nas cidades, e estas, funcionando, como uma infraestrutura para o desenvolvimento do seu capital humano, a saber:

A knowledge city is purposely designed for encouraging and nourishing the collective knowledge, i.e. intellectual capital, as capabilities to shape efficient and sustainable actions of welfare over time. The city can be seen as the structural capital surrounding the human capital but also the relational capital connecting the human capital with the structural capital to give a higher value adding for the knowledge worker (EDVINSON, 2006).

Dimensão Tecnológica (*Technology Dimension*):

- Cidade digital (*Digital City*): segundo Rezende (2012), uma *digital city* é o ambiente digital em rede que interconecta seus serviços públicos, seus setores, pessoas e comunidades gerando dados e informações sobre si mesmo, permitindo que estes agentes interajam entre si em uma realidade virtual, tendo um dos benefícios o encurtamento de distâncias e melhor aproveitamento das áreas urbanas. Faz parte do conceito de *smart cities*.
- *Intelligent City*: Mais uma variação do termo *smart city*, cuja definição passa pela importância das TICs e do capital humano e social, conforme segue:

The notion of an intelligent city emerges at the crossing of the knowledge society (a society in which knowledge and creativity have great emphasis and intangible, human and social capital are considered the most valuable asset) with the digital city (NAM; PARDO, 2011).

- *Wired City*: conceito visionário da década de 60, definindo uma cidade onde a informação de qualquer natureza seria fornecida via satélites, fibra ótica, televisão dentre outros (DUTTON, 1987).
- *Hybrid City*: conceito que é abrangido pelas *smart cities*. Representa a ideia de uma realidade onde os habitantes e demais entidades das cidades possuem uma contraparte virtual

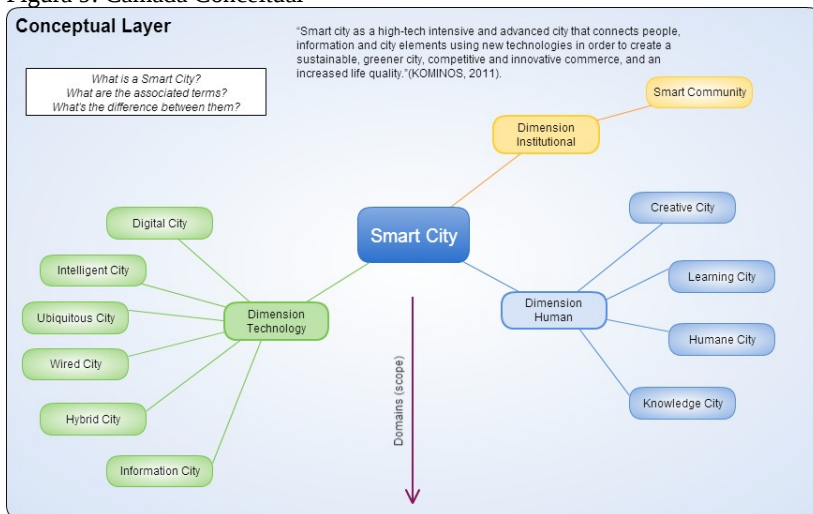
(digital) (NAM; PARDO, 2011, tradução nossa). Exemplo: usuários e carros com cadastro público.

- *Information City*: conforme Nam e Pardo (2011), refere-se ao processo, também incorporado pelas *smart cities* de coleta de dados e informações de comunidades locais e disponibilizá-los via portais Web para o público de uma cidade (ex.: vagas de emprego);
- *Ubiquitous City*: este termo foca na capacidade ubíqua (ou pervasiva) das tecnologias das cidades digitais. Segundo Nam e Pardo (2011), está relacionado à ideia de que os cidadãos podem acessar qualquer serviço, à qualquer hora e em qualquer lugar.

Dimensão Institucional (*Institutional Dimension*):

- *Smart Community*: definido como comunidades, em nível de países, que usam TICS com o objetivo de melhorar qualidade de vida, e onde seus governantes, habitantes e demais instituições trabalham com interesses em comum (NAM; PARDO, 2011 tradução nossa). Nesta dimensão é destacado o papel dos governos como importantes regulamentadores das *smart communities*, e conseqüentemente, das *smart cities*.

Figura 9: Camada Conceitual



Fonte: o autor, baseado em Nam e Pardo (2011) e demais citados.

Tendo definido o termo central *smart city*, fundamentado principalmente em quatro definições encontradas na literatura, foi possível mapear três dimensões importantes deste conceito, além de explicitar qual a visão de *smart cities* que está sendo tratada no Modelo proposto.

A seguir, é descrita a Camadas de Domínios, Indicadores e Atores (*Domain Layer*).

4.3.2 Camada De Domínios, Indicadores E Ações

Após definir o conceito de *smart cities* e suas dimensões conceituais relacionadas, foram então estruturados os mecanismos que regem as *smart cities*, ou seja, seus domínios de atuação. Para isto, foi desenvolvida a Camada de Domínios, Indicadores e Ações (*Domain Layer*). Esta camada apresenta dois grandes grupos de domínios (*hard* e *soft*), conforme Neirotti et al. (2014), que se dividem em seis principais domínios (GIFFINGER et al., 2007; NEIROTTI et al., 2014). Destes seis domínios, conforme diversos autores relatam (ALBINO; BERNARDI; DANGELICO, 2015), cinco são os domínios fundamentais que devem levar ao sexto domínio, chamado de “qualidade de vida” (*smart living*) nas cidades. Os dados de cada domínio são utilizados para gerar indicadores, que servem de insumo para o processo de tomada de decisão nas cidades (ex.: reforçar o controle de perda de água tratada).

Questões de Competência:

- Quais dados coletar (*What data to collect*)?
- Qual o escopo a ser abordado (*What is the scope or subjects to analyse*)?

Apontamentos/Repostas: devem ser coletados dados sobre os seis principais domínios e algumas de suas variações identificadas na literatura sobre *smart cities* (GIFFINGER et al., 2007; NEIROTTI et al., 2014; MINH et al., 2013; ALBINO; BERNARDI; DANGELICO, 2015). Nesta dissertação optou-se por utilizar o termo domínio e subdomínio para referenciar tantos os eixos, de Giffinger et al. (2007), como os domínios (NEIROTTI et al., 2014). Foram também incluídas definições de Albino, Bernardi e Dangelico (2015), Letaifa (2015) e de Minh et al. (2014).

Aos seis domínios apontados por Giffinger et al. (2007) foram adicionados as ações apontadas por Neirotti et al. (2014), cujo trabalho, além da classificação em *hard* e *soft*, traz importantes iniciativas

aplicadas em projetos de *smart cities*. Estas ações foram agrupadas juntamente com os domínios apontados por Giffinger et al. (2007), conforme sua similaridade. São elas: *E-Gov, Document (data) Transparency, Efficient Procurement Management, People Mobility, City Logistics, Human Capital Retaining, Smart Grids, Public Lighting, Waste Management, Water Management, Food and Agriculture Management, Smart Building, Entertainment, Hospitality, Education Facilities, Cultural Heritage Management, Welfare and Social Inclusion, Public Spaces Management* (NEIROTTI et al., 2014).

Após identificar as ações acima, foi necessário classificá-las e reagrupá-las juntamente com os domínios propostos por Giffinger et al. (2007). Os resultados estão descritos a seguir. Os termos foram mantidos em seu idioma original para evitar equívocos em sua tradução e adaptação:

- **Domínio (Hard) Environent (Natural Resources):** *Attractivity of natural conditions, Pollution (Pollution Control), Environmental Protection, Sustainable resource Management (Smart Grids, Public lighting, Green/renewable energies, Water e Waste Management, Food e Agriculture, Buildings);*
- **Domínio (Hard) Mobility (Transport and ICT):** *People Mobility (Local Aecessibility of Public Transport, National Aecessibility e Sustainable Innovate and Safe Transport Systems), Avalability of ICT-infrastrucutre (Mobility-Information and Household-Information), City Logistics;*
- **Domínio (Soft) Economy (Competitiveness/ Innovation):** *Innovative spirit, Enrepreneurship, Economic Image e trademarks, Productivity, Flexibility, International embeddness, Ability to transform;*
- **Domínio (Soft) Governance (Participation e Transparency):** *Participation in decision-making, Public and Social Services, Transparent Governance, Political Strategies and perspectives, E-Government, Document (data) Transparency, Efficient Procurement Management;*
- **Domínio (Soft) People (Social and Human Capital):** *Level of qualification, Affinity to life long learning, Social and ethnic plurality, Flexibility, Creativity, Cosmopolitanism, Participation in public life;*
- **Domínio (Hard/Soft) Living (Quality of life):** *Cultural Facilities (Cultural Heritage Management), Health Conditions, Individual safety, Housing quality, Education facilities,*

Touristic attractivity (Public Spaces Management), Social Cohesion (Welfare and Social Inclusion).

É importante destacar que ao classificar os seis domínios agrupados em dois grandes, grupos *hard* (*mobility, environment, mobility e living*) e *soft* (*economy, government, people e living*), com base na intensidade do impacto das tecnologias baseadas em *big data*, significa dizer que um domínio tende a sofrer um maior ou menor impacto destas tecnologias. Ou seja, o domínio *Environment*, por exemplo, pode sofrer impacto de ações políticas (*soft*), e não apenas de tecnologias e infraestruturas (*hard*).

Indicadores e Ações

Os seis domínios descritos acima formam o escopo de atuação das *smart cities* e podem ser caracterizados e principalmente analisados através de ações e indicadores. Giffinger et al. (2007) elaboram um conjunto de indicadores que visam mensurar resultados em iniciativas de *smart cities*. Eles permitem uma análise sobre o funcionamento das cidades em geral e podem guiar também a construção de novos indicadores para demais projetos de *smart cities*. Além dos indicadores, foram apontadas ações que são comumente associadas às *smart cities* (LOMBARDI et al., 2012; NEIROTTI et al., 2014; LAZAROIU; ROSCIA, 2012), e podem servir como base também para construção de novos indicadores sobre os domínios analisados. Eles seguem descritos abaixo.

Questões de Competência:

- Como extrair informações sobre o mecanismo (os seis domínios) das *smart cities*, para analisar seu funcionamento (*How to extract information about the mechanism/ six domains of smart cities, to analyze its operation*)? Esta se divide em duas questões:
 - Como utilizar os dados coletados sobre os seis domínios (*How to use the collected data*)?
 - Como construir indicadores sobre o desempenho das *smart cities* nestes domínios (*How to build smart cities performance domains indicators*)?

Apontamentos/Respostas: podem ser utilizadas ações e indicadores dos domínios das *smart cities* para analisar seu funcionamento. Para auxiliar esta tarefa, foram levantados indicadores

construídos por Giffinger et al. (2007), baseados em dados disponíveis sobre as cidades europeias, que podem servir como base para construção de novos indicadores para cidades em geral. Alguns destes indicadores devem ser adaptados para a realidade das diferentes cidades fora da Europa (ex.: “*Voters turnout at European elections*” e “*Knowledge about the EU*”).

Além dos indicadores, foram também indicadas ações dentro dos domínios das *smart cities*, que foram principalmente identificadas no trabalho de Neirotti et al. (2014), Minh et al. (2013) e Albino, Bernardi e Dangelico (2015).

As ações e indicadores identificadas foram agrupadas por seus domínios e subdomínios (grupos de fatores), nesta seguinte estrutura: **Domínio, Fator, Indicador** e/ou **Ação**. Ou seja, cada domínio possui um ou mais fatores (ou subdomínios), cada qual com seu grupo de indicadores e ações, conforme segue:

- **Domínio *Smart Economy*:**

- **Fator *Innovative spirit*:**

- **Indicadores:** *R&D expenditure in % of GDP (gross domestic product), Employment rate in knowledge-intensive sectors, Patent applications per inhabitant;*

- **Fator *Entrepreneurship*:**

- **Indicadores:** *Self-employment rate, New businesses registered;*

- **Ações:**

- ***Innovation and entrepreneurship:*** *foster the innovation systems and entrepreneurship in the urban ecosystem (e.g. presence of local incubators);*

- **Fator *Economic image & trademarks*:**

- **Indicadores:** *(Data about) Importance as decision-making centre;*

- **Fator *Productivity*:**

- **Indicadores:** *GDP (gross domestic product) per employed person;*

- **Fator *Flexibility of labour market*:**

- **Indicadores:** *Unemployment rate, Proportion in part-time employment;*

- **Fator *International embeddedness*:**

- **Indicadores:** *Companies with HQ in the city quoted*

- on national stock market, Air transport of passengers, Air transport of freight;
 - **Fator:** *Ability to transform:* sem indicadores ou ações identificadas;
- **Domínio Smart People:**
 - **Fator** *Level of qualification:*
 - **Indicadores:** *Importance as knowledge centre (top research centres, top universities), Population qualified at levels 5-6 ISCED (International Standard Classification of Education ou Classificação Internacional Normalizada da Educação, da UNESCO), Foreign language skills;*
 - **Ações:**
 - **Human Capital Retaining:** *Policies to improve human capital investments, attract and retain new talents, avoiding human capital flight (brain drain);*
 - **Digital Education:** *Extensive Use of modern ICT tools (e.g. interactive whiteboards, e-learning systems) in public schools;*
 - **Fator** *Affinity to life long learning:*
 - **Indicadores:** *Book loans per resident, Participation in life-long-learning in %, Participation in language courses;*
 - **Fator** *Social and ethnic Plurality:*
 - **Indicadores:** *Share of foreigners, Share of nationals born abroad*
 - **Fator** *Flexibility:*
 - **Indicadores:** *Perception of getting a new job;*
 - **Fator** *Creativity:*
 - **Indicadores:** *Share of people working in creative industries;*
 - **Fator** *Cosmopolitanism/Open-mindedness:*
 - **Indicadores:** *Voters turnout at European elections, Immigration-friendly environment (attitude towards immigration), Knowledge about the EU;*
 - **Fator** *Participation in public life:*
 - **Indicadores:** *Voters turnout at city elections, Participation in voluntary work;*

- **Domínio Smart Governance:**
 - **Fator** *Participation in decision-making:*
 - **Indicadores:** *City representatives per resident, Political activity of inhabitants, Importance of politics for inhabitants, Share of female city representatives;*
 - **Fator** *Public and social services:*
 - **Indicadores:** *Expenditure of the municipal per resident in PPS (Prospective Payment System), Share of children in day care, Satisfaction with quality of schools;*
 - **Fator** *Transparent governance:*
 - **Indicadores:** *Satisfaction with transparency of bureaucracy, Satisfaction with fight against corruption;*
 - **Ações:**
 - **E-Government:** *Digitizing the public administration by managing documents and procedures through ICT tools in order to optimise work and offer fast and new services to citizens;*
 - **Document Transparency (ou Open Government Data Initiatives):** *Enabling every citizen to access official documents in a simple way and to take part in the decision processes of a municipality. Decreasing the possibility for authorities of abusing the system for their own interests or hiding relevant information;*
 - **Efficient Procurement Management:** *Allowing the public sector improving procurement procedures and the associated contract management, with the purpose of assuring best value for money without decreasing quality;*
 - **Fator** *Political Strategies and perspectives:* *sem indicadores;*
- **Domínio Smart Mobility:**
 - **Fator** *People Mobility (Local Acessibility of Public Transport, National Acessibility e Sustainable Innovate and Safe Transport Systems):*
 - **Indicadores:** *Public transport network per inhabitant, Satisfaction with access to public transport,*

Satisfaction with quality of public transport; Green mobility share (non-motorized individual traffic), Traffic safety, Use of economical cars;

- **Ações:**

- **Innovate Transport:** Innovative and sustainable ways to provide the transport of people in cities, such as the development of public transport modes and vehicles based on environmental-friendly fuels and propulsion systems, supported by advanced technologies and proactive citizens behaviours;

- **Fator** *Availability of ICT-infrastructure (Mobility-Information and House-Information):*

- **Indicadores (Household-Information):** Computers in households, Broadband internet access in households;

- **Ações (Mobile-Information):** Distributing and using information, both pre-trip and, more importantly, on-trip, with the aim of improving traffic and transport efficiency as well as assuring a high quality travel experience;

- **Fator** *City Logistics:*

- **Ações:** Improving logistics flows in cities by effectively integrating business needs with traffic conditions, geographical, and environmental issues.

- **Domínio Smart Environment:**

- **Fator** *Attractivity of natural conditions:*

- **Indicadores:** Sunshine hours, Green space share;

- **Fator** *Pollution Controll :*

- **Indicadores:** Summer smog (Ozon), Particulate matter, Fatal chronic lower respiratory diseases per inhabitant;

- **Ações:** Controlling emissions and effluents by using different kinds of devices. Stimulating decisions to improve the quality of air, water, and the environment in general;

- **Fator** *Environmental protection:*

- **Indicadores:** Individual efforts on protecting nature, Opinion on nature protection;

- **Ações:** Treat the natural environment as a strategic component for the future;

- **Fator** *Sustainable Resource Management (Smart Grids, Public lighting, Green/renewable Energies, Water e Waste Management, Food e Agriculture):*
 - **Indicadores:** *Efficient use of water (e. g. per capita), Efficient use of electricity (e. g. per capita);*
 - **Ações:**
 - **Smart Grid:** *Implement Electricity networks able to take into account the behaviours of all the connected users in order to efficiently deliver sustainable, economic, and secure electricity supplies. Smart grids should be self-healing and resilient to system anomalies;*
 - **Buildings:**
 - **Smart Building:** *Various systems existing in a building such as electric networks, elevators, fire safety, telecommunication, data processing, and water supply systems. Computer-based systems to control the electrical and mechanical equipment of a building;*
 - *Home and office automation systems (HOS) interconnect electric devices such as heaters, lights, air conditioners, TVs, computers, alarms, and cameras through a communication network, allowing them to be remotely controlled, monitored or accessed from any room in the building, as well as from any location in the world by Internet. They help people to optimize their living style, arrange the day-to-day schedule, secure a high living quality and reduce the energy consumption bill (MINH et al. 2013).*
 - **Facilities Management:** *Cleaning, maintenance, property, leasing, technology, and operating modes associated with facilities in urban areas;*
 - **Public lighting:** *Illumination of public spaces with street lamps that offer different functions, such as air pollution control and Wi-Fi connectivity. Centralised management systems*

that directly communicate with the lampposts can allow reducing maintenance and operating costs, analysing real-time information about weather conditions, and consequently regulating the intensity of light by means of LED technology;

- **Green/renewable Energies:** Exploiting natural resources that are regenerative or inexhaustible, such as heat, water, and wind power;
 - **Water Management:** Analysing and managing the quantity and quality of water throughout the phases of the hydrological cycle and in particular when water is used for agricultural, municipal, and industrial purposes;
 - **Waste Management:** Collecting, recycling, and disposing waste in ways that prevent the negative effects of an incorrect waste management on both people and the environment;
 - **Food and Agriculture:** Wireless sensor networks to manage crop cultivation and know the conditions in which plants are growing. By combining humidity, temperature, and light sensors the risk of frost can be reduced and possible plant diseases or watering requirements based on soil humidity can be detected;
- **Domínio Smart Living:**
 - **Fator Cultural facilities:**
 - **Indicadores:** Cinema attendance per inhabitant, Museums visits per inhabitant, Theatre attendance per inhabitant;
 - **Ações:**
 - **Cultural Heritage Management:** The use of ICT systems (e.g. augmented reality technologies) for delivering new customer experience in enjoying the city's cultural heritage. Use of asset management information systems to handle the maintenance of historical buildings;

- **Fator Health conditions:**
 - **Indicadores:** Life expectancy, Hospital beds per inhabitant, Doctors per inhabitant, Satisfaction with quality of health system;
 - **Ações:** Prevention, diagnosis, and treatment of disease supported by ICT. Assuring efficient facilities and services in the healthcare system;
- **Fator Individual safety:**
 - **Indicadores:** Crime rate, Death rate by assault, Satisfaction with personal safety;
 - **Ações:** Protecting citizens and their possessions through the active involvement of local public organisations, the police force and the citizens themselves. Collecting and monitoring information for crime prevention;
- **Fator Housing quality:**
 - **Indicadores:** Share of housing fulfilling minimal standards, Average living area per inhabitant, Satisfaction with personal housing situation;
- **Fator Education facilities:**
 - **Indicadores:** Students per inhabitant, Satisfaction with access to educational system, Satisfaction with quality of educational system;
- **Fator Touristic attractivity:**
 - **Indicadores:** Importance as tourist location (overnights, sights), Overnights per year per resident;
 - **Ações:**
 - **Public Spaces Management:** Care, maintenance, and active management of public spaces to improve the attractiveness of a city. Solutions to provide information about the main places to visit in a city;
 - **Hospitality:** Ability of a city to accommodate foreign students, tourists, and other non-resident people by offering appropriate solutions to their needs;
- **Fator Social Cohesion:**
 - **Indicadores:** Perception on personal risk of poverty, Poverty rate;

- **Ações:**
 - ***Welfare and Social Inclusion:*** *Social inclusion of various urban residents and social capital in urban development. Improving the quality of life by stimulating social learning and participation, with particular reference to specific categories of citizens such as the elder and disabled.*

Salienta-se aqui que ao fazer análises com os indicadores acima, e com aqueles que forem construídos com base neles, deve-se considerar a periodicidade de produção dos dados (ex.: média anual de consumo de energia elétrica versus média diária).

Os indicadores e ações acima foram adicionados ao Modelo em forma de um conjunto de metadados para a construção de bases de dados, identificação de lacunas de dados, e podem inspirar a construção de novos indicadores. No Capítulo 4.4 (Conjunto de Artefatos Implementados) desta dissertação encontram-se os formulários de metadados construídos para cada domínio, que servem para que o utilizador deste Modelo (ex.: um gestor de uma cidade) possa usá-lo em seus projetos. O uso deste formulário é apresentado no Capítulo 4.5 (Experimento).

Além dos indicadores acima listados, foram também expostos os principais atores (*stakeholders*) influentes no processo de melhorar a qualidade de vida dos habitantes das *smart cities* baseadas em *big data* e *open data*, conforme segue.

Atores

Os principais atores do processo de melhoria da qualidade de vida nas *smart cities* foram identificados na literatura (LOMBARDI et al., 2012; VILAJOSANA et al., 2013; DOBRE; XHAFA, 2013), conforme descrito a seguir. Seus papéis formam a Estrutura do modelo CESM, e mostram como estes atores se relacionam nestas cidades. Por exemplo, a indústria em geral, indústria de TI e desenvolvedores de software auxiliam a gerar o desenvolvimento econômico das cidades, além de promover a inovação tecnológica na geração de novos produtos e serviços (exemplo: Economia Criativa), aumentando assim a competitividade econômica destas cidades. Os fornecedores de *IoT* desenvolvem a infraestrutura das cidades (os sensores das *smart utilities*) agindo diretamente como importantes fontes de tecnologias sustentáveis. A sociedade civil (*civil society*) e o meio acadêmico, representado pelas universidades (*university*), incluindo as demais

instituições de ensino e pesquisa, auxiliam a promover o conhecimento nestas (e sobre estas) cidades, elevando o capital humano e social, e consequentemente seu desenvolvimento como um todo. Todos estes elementos podem ser vistos na Figura 10.

Questões de Competência: Quem são os principais atores nas *smart cities* (*Who are the stakeholders*)? Como eles influenciam as cidades inteligentes (*How do they influence the smart cities*)?

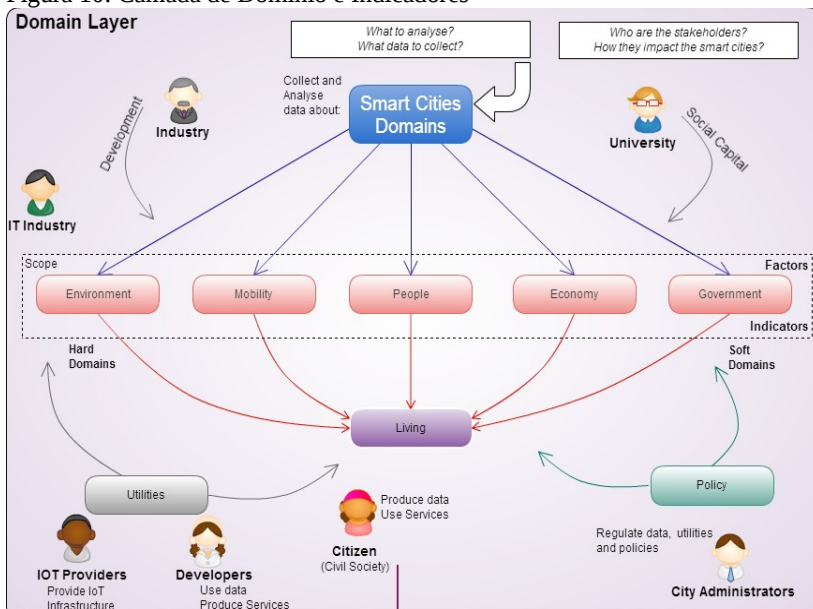
Respostas: seguem abaixo os principais atores identificados, conforme encontrado na literatura e os seus respectivos papéis (LOMBARDI et al., 2012; VILAJOSANA et al., 2013; DOBRE; XHAFA, 2013; DEAKIN, 2011; NEIROTTI et al., 2014):

- *City Administrators*: são os administradores de cidades, pois, principalmente, criam políticas públicas (*policies*) (NEIROTTI et al., 2014); regulamentam o uso dos dados públicos (ex.: *open data*) (VILAJOSANA et al., 2013), e equipam e melhoram os serviços públicos com ferramentas de TICs (*improve utilities*);
- *Civil Society*: representam os cidadãos (Sociedade civil) e as instituições formadas por estes (como ONGs por exemplo). Eles utilizam os serviços públicos inteligentes e os serviços produzidos pelos desenvolvedores de software, produzindo assim dados sobre o seu comportamento na cidade;
- *Industry* (Indústria em geral): geração de riquezas (DEAKIN, 2011) e desenvolvimento através de empregos e inovações, aumentando a competitividade econômica das cidades. Destaca-se aqui o papel da Economia Criativa e as atividades intensivas em conhecimento como fatores estratégicos de sucesso econômico. Deste grupo, destacam-se dois subgrupos:
- *IT Indutry* e *Developers*: representam Indústria de TI e Desenvolvedores de Software em geral. Com base nos dados produzidos pelo uso de TICs nas infraestruturas das cidades, produzem serviços (ex.: aplicativos para dispositivos móveis) que auxiliam a melhorar a vida urbana nestas cidades;
- Fornecedores de tecnologia para *IoT*: fornecem tecnologias (exemplo: sensores) para tornar a infraestrutura das cidades (exemplo: sistemas de abastecimento de água, eletricidade, aterros sanitários) mais eficiente, melhorando o gerenciamento destes recursos;
- *University*: representa os segmentos da sociedade que promovem o conhecimento, como universidades, centros de

pesquisa, incluindo outras instituições de ensino também. Este grupo pode ser chamado de o “mundo acadêmico”. Desenvolvem o capital intelectual (DEAKIN, 2011), através de criação e compartilhamento de conhecimento. Devem também auxiliar no aumento da competitividade econômica e sustentável de uma cidade, através da geração de inovação juntamente com a indústria em geral, além de elevar o capital humano e social das cidades com seu conhecimento produzido;

- Dados (*Data*), Serviços Públicos (*Utilities*), Serviços Desenvolvidos (*Services*) e Políticas Públicas (*Policy*): conforme Dobre e Xhafa (2013) estes itens possuem papéis chaves nas *smart cities*. Através do uso de TICs, os dados são gerados pelas (*smart*) *Utilities*, que servem de insumo para indústria de TI desenvolver seus serviços (*Services*). As políticas públicas (*Policy*), além de regulamentarem o uso e produção destes dados, precisam criar ações para elevar a qualidade dos domínios *soft* (ex.: *smart people*).

Figura 10: Camada de Domínio e Indicadores



Fonte: o autor.

Em seguida, são descritas em detalhes as fontes de dados e suas estruturas, agrupadas na camada *Data Layer* (Camada de Dados),

conforme segue.

4.3.3 Camada De Dados

A Camada de Dados (*Data Layer*) mapeia as principais fontes de dados para os domínios (fatores, ações e indicadores) das *smart cities*, conforme apontado pela literatura, listando e definindo suas principais características, focando *big data* e *open data*, bem como sua proveniência.

Para definir o fenômeno *Big Data* utiliza-se o modelo 5Vs (*velocity, variety, variability, volume* e *value*) (FAN; BIFET, 2012), e também a classificação de *Big Data* da IBM (IBM, 2013), conforme respostas a seguir.

Questões de Competência:

- Quais são as principais fontes de dados (*What are the main data sources*)?
- Dados sobre quais domínios pode-se encontrar nestas fontes (*What smart cities domains data can be found*)?

Respostas:

Foram identificadas as principais fontes, conforme sua natureza de produção: *Human Generated* (Produzidos por Humanos) e *Machine Generated* (Produzidos por Máquinas) (IBM, 2013);

- *Machine Generated*: dados oriundos da infraestrutura de IoT das *smart cities*, (VILAJOSANA et al., 2013; CHEN; MAO, 2014). São produzidos através de redes de sensores, que são implementadas em serviços públicos inteligentes, como redes de abastecimento de água e eletricidade (*smart grids*) por exemplo;
- *Human Generated*: dados produzidos por humanos, através de diversos meios físicos, como computadores pessoais e dispositivos móveis por exemplo. Em relação aos dados produzidos por dispositivos móveis, estes geralmente contêm informações sobre a localização via GPS dos seus usuários. Estes dados tem sido utilizados recentemente para criar soluções urbanas inovadoras, como por exemplo o aplicativo Waze, que é uma aplicação de *smart traffic*, onde um motorista pode interagir com dados em tempo real sobre o trânsito. Conforme afirmado anteriormente, este tipo de aplicativo é considerado pertencente a uma nova geração de programas, baseados em *big data* e na computação perceptível ao contexto (DOBRE; XHAFA, 2013).

As fontes de dados desta camada podem ser agrupados também conforme duas classes de fontes: Redes de Sensores (Serviços Públicos Inteligentes e Dispositivos Móveis) e Web, respectivamente.

As Redes de Sensores (*Network Sensors*) podem ser definidas em dois subgrupos: *Utilities Sensors* (sensores embutidos na infraestrutura das cidades, ex.: redes elétricas) e *Mobile Devices* (dispositivos móveis, que funcionam como sensores utilizados pelos seres humanos, exemplo: *smartphones*).

A Web representa a própria camada Web utilizada na internet. Nela, os dados produzidos sobre a cidades estão disponíveis para utilização através de alguma tecnologia para publicação de dados (ex.: JSON). A internet neste caso, é o seu meio de transmissão. As Redes de Sensores podem ter seus dados publicados via web/internet. Por exemplo: dados abertos governamentais em RDF ou CSV, produzidos por alguma agência governamental, como o setor de Vigilância Sanitária de uma determinada cidade. Dados de sites tradicionais, redes sociais e outras fontes se encaixam neste grupo também. Seguem abaixo maiores detalhes das duas classes *Network Sensors* e Web:

- **Redes de sensores (*Network Sensors*):** rede de sensores (IoT), composta pela infraestrutura estática de sensores (*Sensors of Utilities*) e pelos sensores móveis (*Mobile Devices*), que são compostos pelos dispositivos móveis, conforme seguem:
 - *Utilities Sensors* (sensores de infraestrutura das cidades): dispositivos físicos integrados a instalações das cidades, que reagem a estímulos e geram dados sobre o seu hardware e seu ambiente. Exemplo: *smart grids*, *smart building*, *smart street lighting*. São as principais fontes de dados “*Machine Generated*”. Um exemplo tradicional de uso é o de uma rede (ou sistema) de abastecimento água (*smart water*) com sensores integrada a ela, informando em tempo real sobre eventuais pontos de ruptura na rede, para evitar desperdícios;
 - *Mobile devices* (dispositivos móveis): além dos tradicionais sensores das *smart cities* (em redes de água, calculando a vazão e desperdícios, por exemplo), os atuais dispositivos móveis (*smartphones* e *tablets*, por exemplo) podem atuar também como sensores, coletando e gerando dados sobre o contexto dos cidadãos (sobre seu ambiente, em tempo real). São uma fonte física de dados “*Human Generated*” e “*Machine Generated*”, uma vez que tanto o cidadão pode

produzir dados(ao acessar um aplicativo ou serviço web por exemplo), como os seus dispositivos podem automaticamente produzir dados sobre seu contexto (ex.: localização em tempo real). Estes dispositivos móveis também podem interagir com os sensores estáticos gerando mais informações sobre o contexto. Conforme já mencionado, as informações sobre o contexto de um cidadão em uma determinada situação, possibilitam a criação de uma nova geração de aplicativos perceptíveis ao contexto e baseados em *big data* e *open data*. Isso significa basicamente seres humanos interagindo com seu ambiente, através de dispositivos físicos que funcionam como sensores, gerando dados sobre seu ambiente e tendo seus programas se adaptando em tempo real a estes dados. Esta ideia está relacionada ao conceito de computação sensível ou perceptível ao contexto, onde os computadores podem perceber uma determinada situação e reagir a ela (KIM e ROSSLIN, 2010). Conforme mencionado na fundamentação teórica desta dissertação, os principais dados produzidos por estas fontes, por meio de seus aplicativos, possuem as seguintes características (adaptado e traduzido de Dobre e Xhafa, 2013):

- Dados sobre Mobilidade e localização: dados que informam a localização geográfica de um cidadão/dispositivo móvel (via GPS);
 - Dados *Real-time* de eventos: informações atuais sobre o contexto do usuário. Exemplo: sua localização atual combinada a algum evento que esteja ocorrendo próximo a ele, no mesmo horário (ex.: alguma interrupção em uma rede de água, algum crime, algum acidente de carro que provocou engarrafamento);
 - Dados Interativos com novas fontes: as aplicações que utilizam estes sensores podem descobrir novas fontes de dados próximas aos usuários, definindo seu comportamento. Por exemplo: interação com prédios inteligentes poderiam abrir uma porta ou acionar o elevador;
- **WEB:**
 - Redes Sociais Online - *OSN (online social networks)*: postagens em redes sociais *online*. Segundo relatório da

IBM sobre *smart cities* e mídias sociais (IBM, 2012), as mídias sociais online fornecem um canal para as cidades capturarem e analisarem informações em tempo real sobre seus cidadãos e sua infraestrutura, permitindo que estes reportem problemas;

- Dados Abertos - *Open Data*, conforme segue:
 - *Linked Open Data* (LOD): *web of data* ou dados ligados semanticamente, conforme padrões da Web semântica/*Linked Data* (BERNERS-LEE, 2006; LINKEDDATABOOK, 2015), e abertos, conforme padrões de dados abertos (*Open Data*), que são dados que podem ser acessados, reusados e redistribuídos por qualquer pessoa, e possuem licença aberta (5STARDATA apud BERNERS-LEE; OPENDATAHANDBOOK, 2015; ODI, 2015; OPENDEFINITION, 2015). Diversos setores das cidades podem produzir dados em *open data* e constituir uma fonte de dados. Por exemplo: *Open Science Data*. Destaca-se abaixo os dados abertos governamentais:
 - Dados Abertos Governamentais - OGD - *Open Government Data*: estes dados são a união da tecnologia *Linked Data* (BERNERS-LEE; 2006) com iniciativas de abertura de dados (BIZER; CYGANIAK; GAUS, 2007). Exemplos de iniciativas: <http://www.data.gov>, <http://data.gov.uk/>, conforme ilustrado na fundamentação teórica. Estes dados geralmente são produzidos em plataformas de *e-gov* (governo eletrônico) e disponibilizados em portais de acesso (ex.: <http://dados.gov.br>). Possuem como algumas de suas principais vantagens, o aumento da transparência, um maior controle social para combater a corrupção e o fato de promover o conhecimento e seu compartilhamento entre os atores de uma *smart city*;
- *Public Web Services*: *web services* (ex.: SOAP ou REST) que acessem dados públicos e abertos. São apenas os mesmos dados, sendo disponibilizados de uma forma diferente na Web (via serviços Web). A sua vantagem é facilitar seu acesso diretamente de forma automatizada, através da troca de requisições e mensagens entre sistemas

de informação pela internet;

- Web de Documentos - *Web of Documents*: *web* tradicional. Sites tradicionais ligados através de links sintáticos em hipertexto (HTML) (BERNERS-LEE, 2006). Estes dados podem ser minerados, desde que sejam abertos, porém sua falta de estrutura (dados não estruturados ou semiestruturados geralmente) dificultam sua análise. A definição de estrutura de dados é feita a seguir.

É interessante destacar que para cada fonte de dados utilizada é necessário verificar sua licença de uso, pois nem todo dado do fenômeno *big data* necessariamente será *open data*. Devido a este e outros motivos (transparência, controle social), conforme descrito no capítulo de Fundamentação Teórica, recomenda-se o uso de dados abertos, que não possuam restrições legais para seu uso, reuso e redistribuição, conforme segue nas duas citações abaixo:

Reutilização e Redistribuição: os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados...Abertos (ou Open Data) são dados que podem ser livremente usados, reutilizados e redistribuídos por qualquer pessoa, sujeitos, no máximo, à exigência de atribuição da fonte e compartilhamento pelas mesmas regras (OPENDATAHANDBOOK, 2014);

Livres de licenças: Os dados não estão sujeitos a regulações de direitos autorais, marcas, patentes ou segredo industrial. Restrições razoáveis de privacidade, segurança e controle de acesso podem ser permitidas na forma regulada por estatutos (BRASIL, 2015).

Outro aspecto a ser considerado é a estrutura (ou falta de) dos dados. Caso os dados da fonte utilizada forem abertos (*open data*), eles devem possuir um estrutura mínima que permita seu processamento automático por um computador (facilitando o trabalho do analista), conforme segue:

Data is reasonably structured to allow automated processing. The ability for data to be widely used requires that the data be properly encoded. Free-

form text is not a substitute for tabular and normalized records. Images of text are not a substitute for the text itself. Sufficient documentation on the data format and meanings of normalized data items must be available to users of the data. (OPENGOVDATA, 2015).

A diferença entre dados estruturados, pouco estruturados e não estruturados segue descrita abaixo:

Questões de Competência: em quais formatos são encontrados estes dados (*What are the main content format of the data sources*)?

Respostas: Conforme Sint et al (2009) e IBM (2013), os dados podem ser classificados em estruturados, não estruturados e semiestruturados (*structured, unstructured e semistructured*) :

- *Structured* (Estruturados): dados estruturados através de um modelo (ou esquema) de dados. Exemplos: dados extraídos de sistemas de banco de dados relacionais. Estes dados podem ser acessados geralmente via Web Services, como por exemplo, uma base de endereços e CEPs do correio. Seu esquema varia conforme o sistema que o utiliza (*Variability*). Por exemplo: os dados de um sistema acadêmico podem ser modelados e organizados de maneira diferente em diferentes instituições de ensino. Um estudante para uma instituição de ensino pode estar associado a uma turma, e para outra, a um curso, independente da turma. Essas condições de modelagem devem ser levadas em conta quando se for interpretar (e principalmente integrar) dados em esquemas distintos;
- *Semistructured* (semiestruturados): dados onde o esquema (modelo) de dados está contido nos dados, o que é chamado de modelo “auto-descrito” (BUNEMAN, 1997). Conforme Mello et al. (2000), dados semiestruturados são os dados que apresentam uma representação estrutural heterogênea, não sendo nem completamente não-estruturados nem estritamente tipados. Neste tipo de dado, o esquema de representação está presente juntamente com o dado, exigindo que, ao fazer uma análise sobre ele, é necessário primeiro que sua estrutura seja identificada (MELLO et al., 2000 apud BUN, 1997). O Quadro 8 ilustra a diferença entre dados estruturados e semiestruturados. Exemplos: documentos em formatos específicos (como artigos científicos, currículos vitae), arquivos XML, OWL, JSON e RDF (SINT et al., 2009). Dados comumente encontrados com formatos recomendados pelo

- W3C para a Web semântica, como por exemplo, dados governamentais abertos, em RDF, são uma amostra deste grupo;
- *Unstructured* (não estruturados): produto direto gerado pela comunicação humana, como textos em linguagem natural, e-mails entre outros (OASIS, 2008). São dados sem estrutura definida, auto-descrita (semiestruturados) ou não (estruturados), sem anotações semânticas. Geralmente são analisados com técnicas de *Text Mining* e PLN (processamento de linguagem natural), onde comumente são buscadas palavras chaves para análise de sentimento, identificação de entidades e padrões biológicos dentre outros (MEDELYAN, DIVOLI). Exemplos: textos sem estruturas, trechos de texto em documentos HTML sem anotações semânticas (ex.: conteúdos de sites), áudios, vídeos, imagens. É o caso da fonte de dados *Web of Documents* (sites tradicionais em HTML), descrita anteriormente.

Quadro 8: Dados estruturados, semi e não estruturados

Item de comparação	Dado Não-Estruturado	Dado Estruturado	Dado Semi-estruturado
Tecnologia base (amostra)	Dados binários e caracteres	Tabelas em banco de dados relacional	XML, RDF.
Controle de Transações	Não há	Técnicas maduras implementadas em SGBDs	Técnicas adaptadas de SGBDs e não maduras
Gerenciamento de Versões	Gerenciamento de versão como um todo	Versionamento detalhada, de linhas, colunas	Versionamento não muito comum de triplas ou grafos
Flexibilidade	Alta, não há esquema	Pouco flexível, forte dependência de esquema	Flexível
Escalabilidade	Alta	Baixa. Alterar o	Alta, alterar o

Item de comparação	Dado Não-Estruturado	Dado Estruturado	Dado Semi-estruturado
		esquema é complexo	esquema é simples
Robustez	Não se aplica	Alta, tecnologia estabelecida há mais de 30 anos	Tecnologia nova, ainda não muito difundida

Fonte: adaptado de SINT et al, 2009.

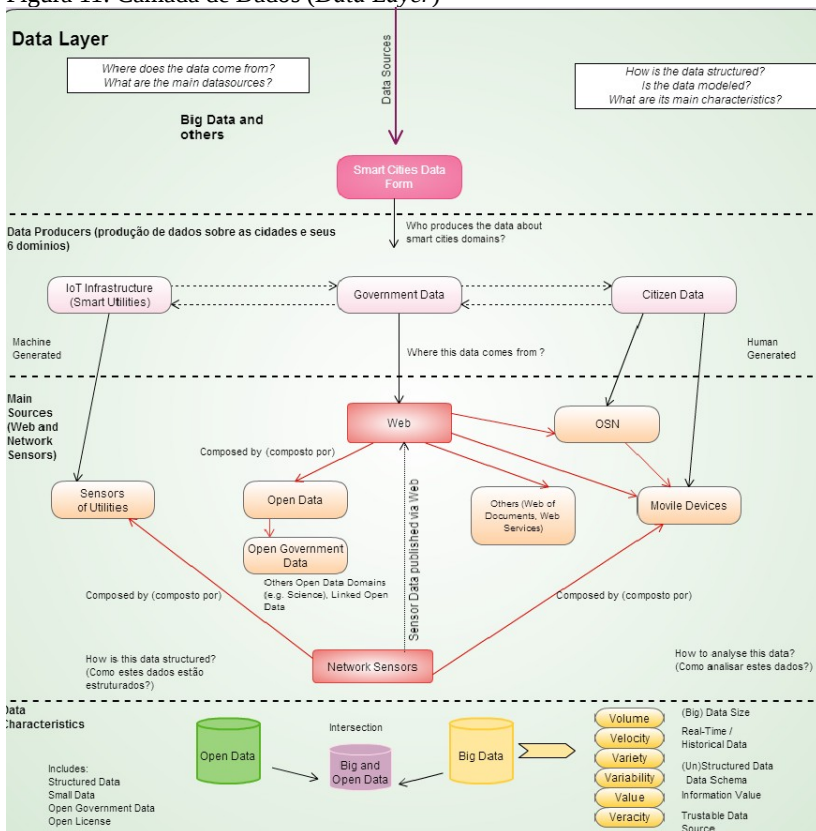
Para os dados coletados nas diversas fontes, estruturados ou não, e para todos os formatos encontrados, é necessário realizar um processo de ETL (extração, tratamento e carga), após sua identificação, para que sejam utilizados (e reutilizados) em sistemas futuros. É importante nesta etapa futura implementar tecnologias que permitam o reuso destas fontes por demais utilizadores, como formatos abertos, como o RDF e vocabulários controlados como o SKOS por exemplo, que provê subsídios para a representação de conhecimento utilizando RDF (W3C, 2016). Como visto na fundamentação sobre *open data*, pesquisadores e outras instituições podem se beneficiar do uso destes dados, produzindo inovação e mais conhecimento sobre as cidades e seus problemas.

Além do modelo 5Vs de Fan e Bifet (2012), adiciona-se o sexto V, de *veracity* (veracidade), que indica o grau de confiança de uma fonte de dados. Este aspecto é muito apontado em relatórios técnicos da IBM, e se deve ao fato de que *big data* inclui dados que podem ser imprecisos, como por exemplo, textos não estruturados, onde seu conteúdo necessita de técnicas processamento de linguagem natural para ser interpretado, podendo gerar equívocos (IBM, 2014; WALKER, 2012). Conforme Alliance (2014, traduzido e adaptado), o aspecto *veracity* corresponde à integridade dos dados, devido a sua (in) precisão, indicando se estes podem ser confiáveis ou não para que uma organização tome decisões cruciais. Neste sentido, as fontes estruturadas e semiestruturadas apresentam um alto grau de confiança, e as não estruturadas um grau mais baixo. A área de Qualidade dos Dados se preocupa em estudar, dentre outras dimensões, a confiabilidade dos dados. Apesar da confiabilidade dos dados depender da sua fonte, rotinas podem ser implementadas para melhorar este item, como por exemplo, a coleta de amostras de dados e verificação da veracidade destas informações no

contexto em que foram produzidos, realizando uma auditoria nas fontes escolhidas.

Continuando com o modelo 5Vs, além dos dados poderem estar em diversos formatos (*variety*) e esquemas (*variability*), eles são produzidos em diferentes periodicidades (*velocity*). Este aspecto é indicado pela questão de competência abaixo:

Figura 11: Camada de Dados (*Data Layer*)



Fonte: o autor.

Questão de Competência: Qual a periodicidade em que eles são produzidos (*What is the data production velocity*)?

Apontamentos: os dados podem ser históricos (acumulados) ou produzidos em tempo real mas serem, por exemplo, disponibilizados em diversos períodos: tempo real (sem diferença de tempo), diariamente,

mensalmente, semestralmente, anualmente. A atualização das bases de dados deve ser realizada conforme sua periodicidade de produção e objetivo do sistema (bases de dados OLTP são mais frequentemente atualizadas que bases OLAP, como em *data warehouses* por exemplo).

O aspecto *Value*, ou valor de negócio, refere-se ao valor das informações que podem estar contidas nestes dados, que deve possibilitar às organizações tomarem decisões que antes (de *big data*) não era possível (FAN; BIFET, 2012, tradução e adaptação nossa). Este item somente é percebido após as análises feitas sobre os dados coletados.

Padrão de Metadados para Proveniência dos Dados

Conforme Quam (2000), o termo meta deriva do grego e significa "denotando uma natureza de uma ordem superior ou tipo mais fundamental", como metalinguagem ou metateoria. Logo, metadados são definidos como atributos que descrevem dados, em um nível de abstração superior aos dados. Metadados podem ser utilizados para descrever a origem, ou proveniência de um certo conjunto de dados.

De acordo com Marins (2008), o conceito de proveniência de dados é definido em padrões de metadados como o Dublin Core, Warwick Framework e norma ISO 19115:2003, sendo apontado resumidamente como a descrição da origem dados. Com o propósito de melhor descrever as fontes dos dados a serem coletados para *smart cities*, foram citadas e utilizadas as questões explicadas por Marins (2008), conforme segue:

- Quando (*When*) o dado foi criado, acessado, modificado? Essa pergunta está ligada a quem acessou ou modificou o dado.
- Onde (*Where*) o dado foi reportado, ou onde está armazenado? Quais são as localizações físicas (no caso de mais de uma cópia)?
- Como (*How*) o dado foi derivado ou transformado? Essa pergunta relaciona-se à qual computação foi aplicada para transformá-lo.
- Quais (*Which*) aplicações, configurações de software ou de ferramentas foram usadas para a criação do dado? Essa questão está ligada às condições do ambiente (MARINS, 2008).

Estas questões descritivas, citadas por Marins (2008), apud Ram (2006), complementam a descrição das fontes de dados. Elas são questões genéricas e podem também ser identificadas no padrão de descrição de metadados Dublin Core. Este padrão pode ser entendido

como o conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos (SOUSA; MELO; VENDRUSCULO, 2000). Seguem os metadados do vocabulário indicado pelo padrão Dublin Core:

Quadro 9: Metadados Dublin Core

Metadado	Definição
Nome	O nome dado para o recurso
Autor ou Criador	A(s) pessoa(s) ou organização(ões) principal(is) responsável(is) pela criação do conteúdo intelectual do recurso. Ex.: autores, no caso de documentos escritos; artistas, fotógrafos ou ilustradores, no caso de recursos visuais.
Palavras-Chave	A essência do conteúdo do recurso. O assunto deve ser expressado por palavras-chave ou frase que descrevam o assunto ou conteúdo do recurso. O uso de vocabulários controlados é encorajado.
Categoria	Atribuição de categoria de assunto que expresse a essência do conteúdo do recurso
Descrição	Uma descrição textual do conteúdo do recurso, incluindo abstracts, no caso de documentos textuais, ou descrição de conteúdo, no caso de recursos visuais
Publicador	A entidade responsável por tornar o recurso disponível na presente forma, tais como uma casa publicadora, um departamento de uma universidade, uma entidade
Colaborador	Pessoa ou organização não especificada no elemento Criador que tenha dado contribuição intelectual significativa para o recurso, mas cuja contribuição é considerada secundária para a pessoa ou instituição especificada no elemento Criador, como, por exemplo, editor, tradutor, ilustrador
Data	A data em que o recurso tornou-se disponível na presente forma
Tipo	A categoria do recurso, como texto, imagem, som, dados, software.
Formato	O formato do dado do recurso, usado para identificar o software e possivelmente o hardware que pode ser

Metadado	Definição
	necessário à exibição ou operação do recurso.
Acesso	Identificador de recursos capturados da rede, como URLs e URNs (quando implementadas). Outros identificadores internacionalmente conhecidos, International Standard Book Number (ISBN), ou outros nomes formais também são candidatos a este elemento.
Identificadores de Acesso	<i>String</i> ou número usado para identificar um recurso, de forma única
Fonte	Informação sobre um segundo recurso do qual o presente recurso é derivado. Embora seja recomendável que elementos contenham informação extraída do presente recurso apenas, o elemento Fonte pode conter uma data, criador, formato, identificador ou outro metadado de um segundo recurso quando este é considerado importante para a identificação do presente recurso.
Idioma	O idioma do conteúdo intelectual do recurso.
Relação	Possibilita relacionamento com outros recursos. A especificação desse elemento visa a fornecer um meio de expressar relacionamentos entre recursos que têm relação formal com outros, mas que existem por si mesmos.
Cobertura	As características espaciais ou temporais do conteúdo intelectual do recurso. Cobertura espacial refere-se a uma região física (por exemplo, celestial); use coordenadas (por exemplo, longitude e latitude) ou nomes de lugares de uma lista controlada. Cobertura temporal refere-se sobre o que é o recurso
Direito Autoral	Uma declaração de direito sobre a propriedade, um identificador que vincula a uma declaração de direito sobre a propriedade, ou um identificador que vincula a um serviço que fornece informação sobre o direito de propriedade do recurso
Contato	Indicação para contato de nome e e-mail da pessoa ou instituição à qual o recurso está vinculado.

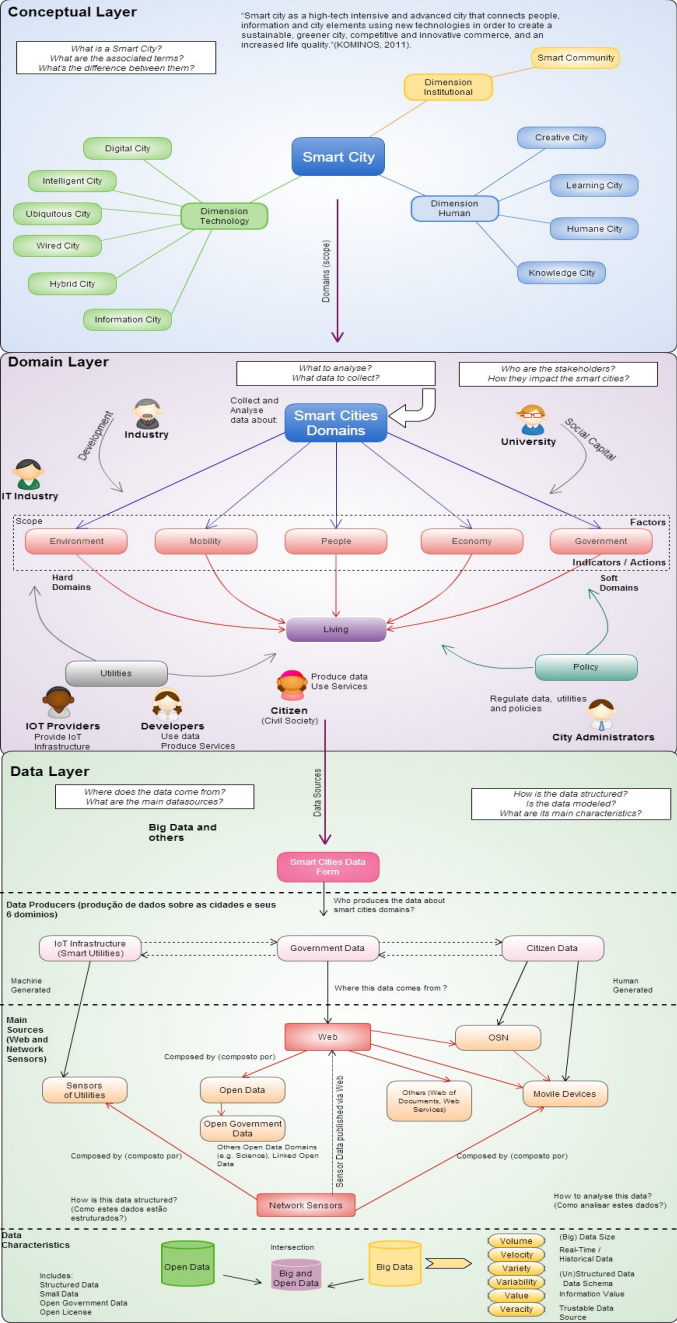
Fonte: o autor, adaptado de Souza; Vendrusculo, Melo (2000).

Em conjunto com as questões genéricas citadas por Marins

(2008), foi aplicado o padrão Dublin Core para catalogar as fontes de dados para as *smart cities*, adaptando e destacando alguns de seus principais atores e elementos. O resultado pode ser visto no capítulo 4.4, que demonstra o conjunto de artefatos implementados.

Finalizando este capítulo, a Figura 12 representa as três camadas construídas que compõem o Modelo conceitual proposto. A seguir, são demonstrados os artefatos implementados para utilização do Modelo proposto.

Figura 12: As três camadas do Modelo Conceitual



4.4 CONJUNTO DE ARTEFATOS IMPLEMENTADOS

Para utilização do Modelo proposto foram implementados sete artefatos em forma de formulários e serem preenchidos. Seus utilizadores devem ser gestores de cidades, com o auxílio de especialistas da área de Ciências da Computação e/ou Sistemas de Informação, como Analistas de Sistemas ou DBAs (Administradores de Banco de Dados) por exemplo. O gestor da cidade por ser o gestor de um departamento, ou de um projeto dedicado a tornar a cidade uma *smart city*, conforme sugerido por Vilajosana et al. (2013). Os especialistas das áreas citadas também têm o conhecimento necessário para utilizar os formulários implementados e podem utilizá-lo diretamente.

Do total dos sete formulários implementados, um formulário identifica, classifica, e registra características e a proveniência das fontes de dados *big data* e *open data*. Para isso, utiliza-se o padrão de metadados Dublin Core, questões genéricas identificadas na literatura e outros metadados adicionados para complementar esta tarefa.

Os outros seis formulários indicam um conjunto de metadados para cada domínio das *smart cities*. Estes metadados descrevem dados, indicadores e ações utilizados em projetos de *smart cities* iniciados na Europa e no mundo, conforme levantado principalmente por Giffinger et al. (2007), Neirotti et al. (2014) e Minh et al. (2013). Eles tem o propósito de guiar a construção de bases de dados, e levantar a necessidade de novos dados a serem produzidos para cada domínio. Após a sugestão destes metadados, para cada domínio devem ser consultados especialistas (Economistas, Engenheiros de Transporte ...), que devem revisar e sugerir novos metadados, conforme necessidades de informação e problemas específicos de cada cidade. Seguem abaixo os formulários implementados.

Quadro 10: Formulário para o mapeamento dos dados por domínio:

Environment

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	<cidade>
Domínio:	<i>Environment</i> (Recursos Naturais)
Identificador:	DO1-E
Fatores Indicados:	

Formulário de metadados do domínio, fatores, indicadores e ações	
Fator: 1	<i>Attractivity of natural conditions</i>
Indicadores:	<i>Sunshine hours, Green space share;</i>
Ações:	--
Fator: 2	<i>Pollution Controll</i>
Indicadores:	<i>Summer smog (Ozon), Particulate matter, Fatal chronic lower respiratory diseases per inhabitant</i>
Ações:	<i>Controlling emissions and effluents by using different kinds of devices. Stimulating decisions to improve the quality of air, water, and the environment in general;</i>
Fator: 3	<i>Environmental protection</i>
Indicadores:	<i>Individual efforts on protecting nature, Opinion on nature protection;</i>
Ações:	<i>Treat the natural environment as a strategic component for the future</i>
Fator: 4	<i>Sustainable Resource Management</i>
Indicadores:	<i>Efficient use of water (e. g. per capita), Efficient use of electricity (e. g. per capita)</i>
Ações:	<p>Smart Grid: <i>Implement Electricity networks able to take into account the behaviours of all the connected users in order to efficiently deliver sustainable, economic, and secure electricity supplies. Smart grids should be self-healing and resilient to system anomalies;</i></p> <p>Smart Building: <i>Various systems existing in a building such as electric networks, elevators, fire safety, telecommunication, data processing, and water supply systems. Computer-based systems to control the electrical and mechanical equipment of a building; Home and office automation systems (HOS) interconnect electric devices such as heaters, lights, air conditioners, TVs, computers, alarms, and cameras through a communication network, allowing them to be remotely controlled, monitored or accessed from any room in the building, as well as from any location in the world by Internet. They help people to optimize their living style, arrange the day-to-day schedule, secure a high living quality and reduce the energy</i></p>

Formulário de metadados do domínio, fatores, indicadores e ações	
	<p>consumption bill;</p> <p>Facilities Management: Cleaning, maintenance, property, leasing, technology, and operating modes associated with facilities in urban areas;</p> <p>Public lighting: Illumination of public spaces with street lamps that offer different functions, such as air pollution control and Wi-Fi connectivity. Centralised management systems that directly communicate with the lampposts can allow reducing maintenance and operating costs, analysing real-time information about weather conditions, and consequently regulating the intensity of light by means of LED technology;</p> <p>Green/renewable Energies: Exploiting natural resources that are regenerative or inexhaustible, such as heat, water, and wind power;</p> <p>Water Management: Analysing and managing the quantity and quality of water throughout the phases of the hydrological cycle and in particular when water is used for agricultural, municipal, and industrial purposes;</p> <p>Waste Management: Collecting, recycling, and disposing waste in ways that prevent the negative effects of an incorrect waste management on both people and the environment;</p> <p>Food and Agriculture: Wireless sensor networks to manage crop cultivation and know the conditions in which plants are growing. By combining humidity, temperature, and light sensors the risk of frost can be reduced and possible plant diseases or watering requirements based on soil humidity can be detected.</p>

Fonte: o autor.

Quadro 11: Formulário para o mapeamento dos dados por domínio: *Mobility*

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	<cidade>
Domínio:	Mobility (Mobilidade, Transporte e TICs)
Identificador:	DO2-M

Formulário de metadados do domínio, fatores, indicadores e ações	
Fatores Indicados:	
Fator: 1	<i>People Mobility</i>
Indicadores:	<i>Public transport network per inhabitant, Satisfaction with access to public transport, Satisfaction with quality of public transport; Green mobility share (non-motorized individual traffic),Traffic safety, Use of economical cars;</i>
Ações:	Innovate Transport: <i>Innovative and sustainable ways to provide the transport of people in cities, such as the development of public transport modes and vehicles based on environmental-friendly fuels and propulsion systems, supported by advanced technologies and proactive citizens behaviours;</i>
Fator: 2	<i>Availability of ICT-infrastructure</i>
Indicadores:	<i>Computers in households, Broadband internet access in households;</i>
Ações:	<i>Distributing and using information, both pre-trip and, more importantly, on-trip, with the aim of improving traffic and transport efficiency as well as assuring a high quality travel experience;</i>
Fator: 3	<i>City Logistics</i>
Indicadores:	--
Ações:	<i>Improving logistics flows in cities by effectively integrating business needs with traffic conditions, geographical, and environmental issues.</i>

Fonte: o autor

Quadro 12: Formulário para o mapeamento dos dados por domínio:

Government

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	
Domínio:	Government (Transparência)
Identificador:	DO3-G

Formulário de metadados do domínio, fatores, indicadores e ações	
Fatores Indicados:	
Fator: 1	<i>Participation in decision-making</i>
Indicadores:	<i>City representatives per resident, Political activity of inhabitants, Importance of politics for inhabitants, Share of female city representatives;</i>
Fator: 2	<i>Public and social services</i>
Indicadores:	<i>Expenditure of the municipal per resident in PPS (Prospective Payment System), Share of children in day care, Satisfaction with quality of schools;</i>
Fator: 3	<i>Transparent governance:</i>
Indicadores:	<i>Satisfaction with transparency of bureaucracy, Satisfaction with fight against corruption;</i>
Ações:	<p>E-Government: Digitizing the public administration by managing documents and procedures through ICT tools in order to optimise work and offer fast and new services to citizens;</p> <p>Document Transparency (ou Open Government Data Initiatives): Enabling every citizen to access official documents in a simple way and to take part in the decision processes of a municipality. Decreasing the possibility for authorities of abusing the system for their own interests or hiding relevant information;</p> <p>Efficient Procurement Management: Allowing the public sector improving procurement procedures and the associated contract management, with the purpose of assuring best value for money without decreasing quality;</p>
Fator: 4	<i>Political Strategies and perspectives</i>
Indicadores:	--

Fonte: o autor.

Quadro 13: Formulário para o mapeamento dos dados por domínio: *Economy*

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	<cidade>

Formulário de metadados do domínio, fatores, indicadores e ações	
Domínio:	<i>Economy (Competitividade e Economia Criativa)</i>
Identificador :	DO4-E
Fatores Indicados:	
Fator: 1	<i>Innovative spirit</i>
Indicadores:	<i>R&D expenditure in % of GDP (gross domestic product), Employment rate in knowledge-intensive sectors, Patent applications per inhabitant;</i>
Fator: 2	<i>Entrepreneurship</i>
Indicadores:	<i>Self-employment rate, New businesses registered;</i>
Ações:	<i>Foster the innovation systems and entrepreneurship in the urban ecosystem (e.g. presence of local incubators);</i>
Fator: 3	<i>Economic image & trademarks:</i>
Indicadores:	<i>(Data about) Importance as decision-making centre;</i>
Fator: 4	<i>Productvity</i>
Indicadores:	<i>GDP (gross domestic product) per employed person</i>
Fator: 5	<i>Economic image & trademarks</i>
Indicadores:	<i>GDP (gross domestic product) per employed person;</i>
Fator: 6	<i>Flexibility of labour market</i>
Indicadores:	<i>Unemployment rate, Proportion in part-time employment;</i>
Fator: 7	<i>International embeddedness</i>
Indicadores:	<i>Companies with HQ in the city quoted on national stock market, Air transport of passengers, Air transport of freight;</i>
Fator: 8	<i>Ability to transform</i>
Indicadores:	--

Fonte: o autor.

Quadro 14: Formulário para o mapeamento dos dados por domínio: *People*

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	<cidade>
Domínio:	<i>People</i> (Capital Humano e Social)
Identificador:	DO5-P
Fatores Indicados:	
Fator: 1	<i>Level of qualification</i>
Indicadores:	<i>Importance as knowledge centre (top research centres, top universities), Population qualified at levels 5-6 ISCED (International Standard Classification of Education ou Classificação Internacional Normalizada da Educação, da UNESCO), Foreign language skills;</i>
Ações:	<i>Human Capital Retaining: Policies to improve human capital investments, attract and retain new talents, avoiding human capital flight (brain drain); Digital Education: Extensive Use of modern ICT tools (e.g. interactive whiteboards, e-learning systems) in public schools;</i>
Fator: 2	<i>Affinity to life long learning</i>
Indicadores:	<i>Book loans per resident, Participation in life-long-learning in %, Participation in language courses;</i>
Fator: 3	<i>Social and ethnic Plurality:</i>
Indicadores:	<i>Share of foreigners, Share of nationals born abroad</i>
Fator: 4	<i>Flexibility</i>
Indicadores:	<i>Perception of getting a new job;</i>
Fator: 5	<i>Creativity</i>
Indicadores:	<i>Share of people working in creative industries;</i>
Fator: 6	<i>Cosmopolitanism/Open-mindedness</i>
Indicadores:	<i>Voters turnout at European elections, Immigration-friendly environment (attitude towards immigration), Knowledge about the EU;</i>
Fator: 7	<i>Participation in public life:</i>

Formulário de metadados do domínio, fatores, indicadores e ações	
Indicadores:	<i>Voters turnout at city elections, Participation in voluntary work;</i>

Fonte: o autor.

Quadro 15: Formulário para o mapeamento dos dados por domínio: *Living*

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	<cidade>
Domínio:	<i>Living (Qualidade de Vida)</i>
Identificador:	DO6-L
Fatores Indicados:	
Fator: 1	<i>Cultural facilities</i>
Indicadores:	<i>Cinema attendance per inhabitant, Museums visits per inhabitant, Theatre attendance per inhabitant;</i>
Ações:	Cultural Heritage Management: <i>The use of ICT systems (e.g. augmented reality technologies) for delivering new customer experience in enjoying the city's cultural heritage. Use of asset management information systems to handle the maintenance of historical buildings;</i>
Fator: 2	<i>Health conditions</i>
Indicadores:	<i>Life expectancy, Hospital beds per inhabitant, Doctors per inhabitant, Satisfaction with quality of health system;</i>
Ações:	<i>Prevention, diagnosis, and treatment of disease supported by ICT. Assuring efficient facilities and services in the healthcare system;</i>
Fator: 3	<i>Individual Safety</i>
Indicadores:	<i>Crime rate, Death rate by assault, Satisfaction with personal safety;</i>
Ações:	<i>Protecting citizens and their possessions through the active involvement of local public organisations, the police force and the citizens themselves. Collecting and monitoring information for crime prevention;</i>
Fator: 4	<i>Housing quality</i>

Formulário de metadados do domínio, fatores, indicadores e ações	
Indicadores:	<i>Share of housing fulfilling minimal standards, Average living area per inhabitant, Satisfaction with personal housing situation;</i>
Fator: 5	<i>Education facilities</i>
Indicadores:	<i>Students per inhabitant, Satisfaction with access to educational system, Satisfaction with quality of educational system;</i>
Fator: 6	<i>Touristic attractivity</i>
Indicadores:	<i>Importance as tourist location (overnights, sights), Overnights per year per resident;</i>
Ações:	<p>Public Spaces Management: <i>Care, maintenance, and active management of public spaces to improve the attractiveness of a city. Solutions to provide information about the main places to visit in a city;</i></p> <p>Hospitality: <i>Ability of a city to accommodate foreign students, tourists, and other non-resident people by offering appropriate solutions to their needs;</i></p>
Fator: 7	<i>Social Cohesion</i>
Indicadores:	<i>Perception on personal risk of poverty, Poverty rate;</i>
Ações:	<i>Welfare and Social Inclusion: Social inclusion of various urban residents and social capital in urban development. Improving the quality of life by stimulating social learning and participation, with particular reference to specific categories of citizens such as the elder and disabled.</i>

Fonte: o autor.

Para o formulário de mapeamento de fontes de dados *big data* e *open data* para *smart cities* aplicou-se o padrão de metadados Dublin Core e outros metadados, conforme segue descrito e demonstrado nos Quadros 16 e 17.

Quadro 16: Dublin Core Aplicado a *Smart Cities*

Metadado	Definição	Metadado Smart City
Nome	O nome dado para o recurso	Nome da Fonte

Metadado	Definição	Metadado <i>Smart City</i>
Autor ou Criador	A(s) pessoa(s) ou organização(ões) principal(is) responsável(is) pela criação do conteúdo intelectual do recurso. Ex.: autores, no caso de documentos escritos; artistas, fotógrafos ou ilustradores, no caso de recursos visuais.	Autor (es)
Palavras-Chave	A essência do conteúdo do recurso. O assunto deve ser expressado por palavras-chave ou frase que descrevam o assunto ou conteúdo do recurso. O uso de vocabulários controlados é encorajado.	Principais Fatores: ex.: <i>Attractivity of natural conditions</i> Principais Indicadores, Ações: ex.: <i>Green space share</i> .
Categoria	Atribuição de categoria de assunto que expresse a essência do conteúdo do recurso	Domínio atendido. Ex.: <i>Smart Environment, Smart People ...</i>
Descrição	Uma descrição textual do conteúdo do recurso, incluindo <i>abstracts</i> , no caso de documentos textuais, ou descrição de conteúdo, no caso de recursos visuais	Descrição da Fonte
Publicador	A entidade responsável por tornar o recurso disponível na presente forma, tais como uma casa publicadora, um departamento de uma universidade, uma entidade	Publicador
Colaborador	Pessoa ou organização não especificada no elemento Criador que tenha dado contribuição intelectual significativa para o recurso, mas cuja contribuição é considerada secundária para a pessoa ou instituição especificada no elemento	Colaborador (editor, diagramador, tradutor ...)

Metadado	Definição	Metadado Smart City
	Criador, como, por exemplo, editor, tradutor, ilustrador	
Data	A data em que o recurso tornou-se disponível na presente forma	Data de Disponibilização do dado
Tipo	A categoria do recurso, como texto, imagem, som, dados, software.	Tipo
Formato	O formato do dado do recurso, usado para identificar o software e possivelmente o hardware que pode ser necessário à exibição ou operação do recurso.	Formato
Acesso	Identificador de recursos capturados da rede, como URLs e URNs (quando implementadas). Outros identificadores internacionalmente conhecidos, International Standard Book Number (ISBN), ou outros nomes formais também são candidatos a este elemento.	URL
Identificadores de Acesso	<i>String</i> ou número usado para identificar um recurso, de forma única	Identificação única: ISSN, ISBN, URI (URL) ...
Fonte	Informação sobre um segundo recurso do qual o presente recurso é derivado. Embora seja recomendável que elementos contenham informação extraída do presente recurso apenas, o elemento Fonte pode conter uma data, criador, formato, identificador ou outro metadado de um segundo recurso quando este é considerado importante para a identificação do presente recurso.	Referência ou <i>link</i> relevante contido no dado
Idioma	O idioma do conteúdo intelectual do recurso.	Idioma
Relação	Possibilita relacionamento com outros	Utilizado no

Metadado	Definição	Metadado <i>Smart City</i>
	recursos. A especificação desse elemento visa a fornecer um meio de expressar relacionamentos entre recursos que têm relação formal com outros, mas que existem por si mesmos.	item Fonte
Cobertura	As características espaciais ou temporais do conteúdo intelectual do recurso. Cobertura espacial refere-se a uma região física (por exemplo, celestial); use coordenadas (por exemplo, longitude e latitude) ou nomes de lugares de uma lista controlada. Cobertura temporal refere-se sobre o que é o recurso.	Nome da Cidade Periodicidade de Produção (tempo-real, diário, mensal)
Direito Autoral	Uma declaração de direito sobre a propriedade, um identificador que vincula a uma declaração de direito sobre a propriedade, ou um identificador que vincula a um serviço que fornece informação sobre o direito de propriedade do recurso	Licença de Uso (Open data ou não)
Contato	Indicação para contato de nome e e-mail da pessoa ou instituição à qual o recurso está vinculado.	Contato

Fonte: o autor.

Ainda ao padrão Dublin Core foram adicionadas e adaptadas as questões genéricas de Marins (2008), apud Ram (2006), que tratam de questões computacionais relacionadas ao tratamento e manipulação dos dados utilizados (“Como o dado foi derivado ou transformado” e “Quais são as configurações de software utilizadas”).

-Dado bruto ou transformado com processo de ETL?

-Quais ferramentas de software foram utilizadas para produzir e/ou publicar os dados?

Além destes metadados, foram adicionados os metadados baseados nas características de fontes de dados *big data* e *open data*: “natureza dos dados” (dados produzidos por humanos ou automaticamente por máquinas), “plataforma de coleta dos dados”

(sensores, web), “detalhamento da plataforma” (dispositivos móveis, sensores em serviços públicos, redes sociais), “formas de disponibilização” (dado aberto, postagem em redes sociais), “departamento governamental” (no caso de DGA – Dados Governamentais Abertos), “estrutura do dado” (estruturado, semi ou não estruturado), “periodicidade de produção” (em tempo real, diária, mensal) e “tamanho” (KB, MB, GB, TB, PB, etc). O metadado “Departamento Governamental” também é utilizado no padrão de metadados para publicação de dados *open data* do governo americano (PROJECT OPEN DATA, 2015).

Os novos metadados adicionados, a aplicação do Dublin Core e a adaptação das perguntas genéricas seguem implementadas no formulário de mapeamento de fontes de dados, conforme Quadro 17.

Quadro 17: Formulário para proveniência das fontes de dados *big data* e *open data* para *smart cities*

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
Cidade:	<cidade>
Domínio	() <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador:	<Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i> >
Fontes de Dados:	
Fonte 1:	<nome da fonte>
Descrição da Fonte:	<descrição textual do conteúdo do recurso>
Entidade Responsável pelos dados:	<publicadores dos dados. ex.: universidades, organizações, órgãos públicos>
Contato	<nome, e-mail, telefone do responsável pelos dados>
Autor dos dados	<autores responsáveis pelo conteúdo intelectual dos dados, quando houver indicação. ex.: autores de relatórios técnicos>
Colaboradores	<pessoas ou organizações que tenha contribuído intelectualmente para o dado, quando informado.ex.: editor, tradutor, ilustrador>

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

Data de disponibilização dos dados:	<data de publicação do dado>
Acesso ao dado:	<forma de acesso ao dado ex.: URL do repositório, DOI, ISSN>
Identificador:	< <i>String</i> ou código identificador da fonte de dados.ex.: ISBN, URI (URL) >
Fonte:	<referência ou <i>link</i> relevante contido no dado>
Idioma:	<idioma dos dados>
Relação	<relação específica com outro recursos ou fontes de dados > Identificador do Recurso: Nome do recurso/fonte: URL: Relação: <input type="checkbox"/> IsPartOf HasPart <input type="checkbox"/> IsVersionOf HasVersion <input type="checkbox"/> IsFormatOf HasFormat <input type="checkbox"/> References IsReferencedBy <input type="checkbox"/> IsBasedOn IsBasisFor <input type="checkbox"/> Requires IsRequiredBy
Cobertura Temporal do dado:	<ex. dados sobre vazamento de rede de água no período de 2010 a 2012 >
Dado bruto ou transformado ?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte>
Quais softwares são necessários?	<ferramentas de software que foram utilizadas para produzir e/ou publicar os dados>
Dados do domínio da <i>smart city</i>	
Principais fatores atendidos pela fonte:	<conforme retorno do domínio escolhido>
Principais Indicadores e ações atendidos pela	<conforme retorno do domínio escolhido, agrupado por fator>

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
fonte:	
Características <i>big data</i> e <i>open data</i> (<i>Volume, Variety, Velocity, Variability</i> , Licença <i>Open Data</i> , formato ...):	
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...>
Natureza dos dados:	<input type="checkbox"/> Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) <input type="checkbox"/> Produzido por Humanos (ex.: <i>posts</i> em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação) <input type="checkbox"/> Não Informado
Plataforma da Fonte de dados (meio onde foi coletado):	<input type="checkbox"/> Dados coletados de Sensores <input type="checkbox"/> Web <input type="checkbox"/> Outro. Especificar: _____ _____
Detalhamento da plataforma (meio onde foi produzido):	<input type="checkbox"/> Serviços Públicos <input type="checkbox"/> Dispositivos Móveis <input type="checkbox"/> Sites <input type="checkbox"/> Outro. Especificar: _____ _____
Forma de disponibilização:	<input type="checkbox"/> Dados Abertos <input type="checkbox"/> <i>Web Services</i> <input type="checkbox"/> Postagens em Redes Sociais <input type="checkbox"/> Outra. Especificar: _____ _____
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.: coordenadas geográficas) da fonte de dados? <input type="checkbox"/> Sim <input type="checkbox"/> Não
DGA - Agência	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito, Secretaria de Saúde) :
Licença	Possui licença Aberta (Dado Aberto) ? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/>

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

Open Data	Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? () Sim () Não () Não Informado Especificar: _____
Estrutura do Dado:	() Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) () Semiestruturados (ex.: RDF, XML, onde o esquema do dado está juntamente com o dado) () Não-estruturados (sem esquema de dados, ex.: textos, sites em HTML)
Periodicidade de Produção:	<Ex.: tempo-real, diária, mensal, anual>:
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1 TB):

Fonte: o autor.

A seguir, é realizada uma verificação do Modelo através do uso destes formulários em um cenário de uso.

4.5 EXPERIMENTO: VERIFICAÇÃO DO MODELO EM UM CENÁRIO DE USO

Com o propósito de verificar a utilização do Modelo proposto nesta dissertação, foi escolhido um dos domínios das *smart cities* para analisar os resultados. Foi escolhido o domínio *Environment* (Recursos Naturais) e a cidade de São Paulo para executar a verificação. O usuário deve ser um gestor de cidades com o auxílio de um analista de sistemas, DBA (administrador de bancos de dados), ou outro profissional de áreas afins, e ao final de seu uso, especialistas de cada domínio (mobilidade urbana, meio ambiente, saúde, segurança etc) devem ser consultados. A verificação é feita em dois passos, conforme segue.

Passo 1: primeiramente, o usuário define qual domínio ele quer analisar. Após isso, o Modelo sugere um conjunto de metadados a serem coletados (e/ou construídos, caso a cidade não disponha destes dados). Estes metadados são agrupados em fatores, indicadores e ações. Estes

fatores também servem como base para a construção de novos indicadores e ações, conforme a necessidade do seu usuário (Quadro 18). Para cada domínio, existem de 1 até N fatores; e cada fator possui de 0 (zero) a N indicadores e ou ações.

Quadro 18: Formulário para o mapeamento dos dados por domínio:

Environment

Formulário de metadados do domínio, fatores, indicadores e ações	
Cidade:	São Paulo
Domínio:	<i>Environment (Recursos Naturais)</i>
Identificador:	D01-E
Fatores Indicados:	
Fator: 1	<i>Attractivity of natural conditions</i>
Indicadores:	<i>Sunshine hours, Green space share;</i>
Ações:	--
Fator: 2	<i>Pollution Controll</i>
Indicadores:	<i>Summer smog (Ozon), Particulate matter, Fatal chronic lower respiratory diseases per inhabitant</i>
Ações:	<i>Controlling emissions and effluents by using different kinds of devices. Stimulating decisions to improve the quality of air, water, and the environment in general;</i>
Fator: 3	<i>Environmental protection</i>
Indicadores:	<i>Individual efforts on protecting nature, Opinion on nature protection;</i>
Ações:	<i>Treat the natural environment as a strategic component for the future</i>
Fator: 4	<i>Sustainable Resource Management</i>
Indicadores:	<i>Efficient use of water (e. g. per capita), Efficient use of electricity (e. g. per capita)</i>
Ações:	Smart Grid: <i>Implement Electricity networks able to take into account the behaviours of all the connected users in order to efficiently deliver sustainable, economic, and secure electricity supplies. Smart grids should be self-</i>

Formulário de metadados do domínio, fatores, indicadores e ações

healing and resilient to system anomalies;

Smart Building: Various systems existing in a building such as electric networks, elevators, fire safety, telecommunication, data processing, and water supply systems. Computer-based systems to control the electrical and mechanical equipment of a building; Home and office automation systems (HOS) interconnect electric devices such as heaters, lights, air conditioners, TVs, computers, alarms, and cameras through a communication network, allowing them to be remotely controlled, monitored or accessed from any room in the building, as well as from any location in the world by Internet. They help people to optimize their living style, arrange the day-to-day schedule, secure a high living quality and reduce the energy consumption bill;

Facilities Management: Cleaning, maintenance, property, leasing, technology, and operating modes associated with facilities in urban areas;

Public lighting: Illumination of public spaces with street lamps that offer different functions, such as air pollution control and Wi-Fi connectivity. Centralised management systems that directly communicate with the lampposts can allow reducing maintenance and operating costs, analysing real-time information about weather conditions, and consequently regulating the intensity of light by means of LED technology;

Green/renewable Energies: Exploiting natural resources that are regenerative or inexhaustible, such as heat, water, and wind power;

Water Management: Analysing and managing the quantity and quality of water throughout the phases of the hydrological cycle and in particular when water is used for agricultural, municipal, and industrial purposes;

Waste Management: Collecting, recycling, and disposing waste in ways that prevent the negative effects of an incorrect waste management on both people and the environment;

Food and Agriculture: Wireless sensor networks to manage crop cultivation and know the conditions in which plants are growing. By combining humidity, temperature, and light sensors the risk of frost can be reduced and possible plant

Formulário de metadados do domínio, fatores, indicadores e ações	
	<i>diseases or watering requirements based on soil humidity can be detected.</i>

Fonte: o autor.

Os fatores, indicadores e ações indicados pelo Modelo funcionam como metadados genéricos, que apontam quais dados devem ser analisados pelo seu utilizador. No caso específico das ações, os metadados estão no gerúndio pois apontam iniciativas utilizadas em *smart cities* ao redor do mundo, as quais se implementadas, podem gerar dados. Com estes metadados, um gestor de cidades pode procurar por fontes, ou basear-se neste apontamento e construir novos fatores, indicadores e ações.

Considerando o retorno do Passo 1, escolheu-se um subconjunto de metadados a serem analisados pelo utilizador. Para cada metadado buscou-se por fontes de dados e aplicou-se o Passo 2, a seguir. Os metadados escolhidos foram seguintes:

- Fator *Attractivity of Natural Conditions* (Atratividade das Condições Naturais):
 - Indicador: *Green space share* (Área Verde Por Habitante);
- Fator *Sustainable Resource Management/Water e Waste Management* (Gereciamento Sustentável de Recursos Naturais, foco em Água e Resíduos Sólidos):
 - Indicadores:
 - *Efficient use of water use per GDP* (Uso eficiente de água);
 - *Efficient use of electricity use per GDP* (Uso eficiente de eletricidade);
- Fator *Environmental Protection* (Proteção Ambiental):
 - Indicadores:
 - *Individual efforts on protecting nature protection* (Iniciativas dos cidadãos para preservação do meio ambiente);
 - *Opinion on nature protection* (opinião sobre a preservação ambiental).

Passo 2: mapear fontes de dados para os metadados indicados no passo anterior. Para isso, foi aplicado o formulário de características e proveniência de fontes de dados para cada fonte de dado identificada.

A cidade escolhida foi São Paulo, pela sua maior disponibilidade

de dados. As fontes de dados identificadas para cada fator seguem descritas nos formulários aplicados.

Para os dados do indicador *Efficient use of electricity use per GDP* (Uso eficiente de eletricidade), foi encontrado o dado “consumo total de eletricidade per capita”, da fonte ELETROPAULO (responsável pela distribuição de energia na cidade de São Paulo).

Como fonte para o indicador de “Uso Eficiente de Água” foram utilizados dados sobre perda de água tratada (percentagem do total de água tratada), cuja fonte é a SABESP (Companhia de Saneamento Básico do Estado de São Paulo).

Para o fator *Attractivity of Natural Conditions* (indicador *Green space share*-Área Verde Por Habitante) foram utilizados dados da fonte SVMA – Secretaria do Verde e Meio Ambiente da cidade de São Paulo. Outra fonte de dados foi utilizada para compor este fator, focando em resíduos sólidos (*Waste Management*) foi a SES Secretaria Municipal de Serviços. Para isso, foi adicionado o novo indicador “Reciclagem de Resíduos Sólidos (percentagem)”, pois este dado foi encontrado para uso no Programa Cidades Sustentáveis, portal de agrupamento de dados governamentais do programa Cidades Sustentáveis (EOEKE, 2015), de onde os dados das quatro fontes usadas até agora (SVMA, SES, SABESP e ELETROPAULO) foram obtidas.

Para compor o fator *Environmental Protection*, mais precisamente os indicadores: *Individual efforts on protecting nature protection* (Iniciativas dos cidadãos para preservação do meio ambiente) e *Opinion on nature protection* (opinião sobre a preservação ambiental) não foram encontrados dados para a cidade de São Paulo. Logo, foi utilizado um relatório em PDF que apresenta os resultados de uma pesquisa de opinião realizada com a população brasileira sobre o tema “desenvolvimento sustentável”, elaborado pelo Ministério do Meio Ambiente do Governo Federal.

Para somar ao indicador *Opinion on nature protection* (opinião sobre a preservação ambiental), foram também coletados dados da rede social Twitter. Estes *posts* podem ser minerados com técnicas de extração de conhecimento em bases textuais e análise de sentimento por exemplo, para avaliar a opinião emitida pelos usuários, como é visto em muitas pesquisas atuais e aplicativos comerciais. No caso do Twitter, o filtro geográfico por cidade torna pequeno o resultado, apenas 8 postagens, devido à falta desta informação na base de dados. Segundo Reino (2015), mesmo sendo um percentual de 77% o número de *tweets* postados na rede através de dispositivos móveis, apenas uma pequena parte é georreferenciada. O Twitter fornece acesso via sua API aos dados

georreferenciados, porém o serviço fornece apenas os *tweets* postados nos últimos sete dias (TWITTER, 2015). Para fim de demonstração de utilização deste formulário em dados não estruturados, foi utilizado o protocolo HTTP GET para filtrar os *tweets* por região e data (desde 2009 até 2016), postados próximo a São Paulo, através dos parâmetros *since*, *until* e *within* respectivamente, oferecidos pelo portal de buscas avançadas do Twitter.

Quadro 19: Fonte de Dados ELETROPAULO – Consumo de Energia per capita

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
Cidade:	São Paulo
Domínio	(x) <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador:	DO1-E <Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i>>
Fontes de Dados:	
Fonte 1:	ELETROPAULO
Descrição da Fonte:	Dados sobre consumo de energia per capita na cidade de São Paulo
Entidade Responsável pelos dados:	ELETROPAULO Programa Cidades Sustentáveis (Publicador dos dados).
Contato	https://www.aeseletropaulo.com.br/Paginas/aes-eletropaulo.aspx
Autor dos dados	--
Colaboradores	--
Data de disponibilização dos dados:	--
Acesso ao dado:	http://www.cidadessustentaveis.org.br

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
Identificador:	http://indicadores.cidadessustentaveis.org.br/br/SP/sao-paulo/indicadores.xls
Fonte:	--
Idioma:	Português
Relação	<relação específica com outro recursos ou fontes de dados > Identificador do Recurso: Nome do recurso/fonte: URL: Relação: <input type="checkbox"/> IsPartOf HasPart <input type="checkbox"/> IsVersionOf HasVersion <input type="checkbox"/> IsFormatOf HasFormat <input type="checkbox"/> References IsReferencedBy <input type="checkbox"/> IsBasedOn IsBasisFor <input type="checkbox"/> Requires IsRequiredBy
Cobertura Temporal do dado:	2009-2009
Dado bruto ou transformado ?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte> Dado Tratado
Quais softwares são necessários?	<ferramentas de software que foram utilizadas para produzir e ou publicar os dados> Editor de planilha eletrônica, como LibreOffice Calc, Microsoft Excel.
Dados do domínio da <i>smart city</i>	
Principais fatores atendidos pela fonte:	<i>Sustainable Resource Management/Water e Waste Management</i>
Principais Indicadores e ações atendidos pela fonte:	<i>Efficient use of electricity use per GDP (Uso eficiente de eletricidade)</i>
Características <i>big data</i> e <i>open data</i> (Volume, Variety, Velocity, Variability, Licença <i>Open Data</i> , formato ...):	

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...> Tabular
Natureza dos dados:	() Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) (x) Produzido por Humanos (ex.: <i>posts</i> em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação) () Não Informado
Plataforma da Fonte de dados (meio onde foi coletado):	() Dados coletados de Sensores (x) Web () Outro. Especificar: _____
Detalhamento da plataforma (meio onde foi produzido):	() Serviços Públicos () Dispositivos Móveis (x) Sites () Outro. Especificar: _____
Forma disponibilização:	() Dados Abertos () <i>Web Services</i> () Postagens em Redes Sociais (x) Outra. Especificar: Planilha do Microsoft Office Excel
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.: coordenadas geográficas) da fonte de dados? () Sim (x) Não
DGA	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito): Secretaria do Meio Ambiente da Prefeitura de São Paulo
Licença <i>Open Data</i>	Possui licença Aberta (Dado Aberto) ? () Sim () Não (x) Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? () Sim () Não (x) Não Informado

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

	Especificar: _____ _____
Estrutura do Dado:	(x) Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) () Semiestruturados (ex.: RDF, XML, onde o esquema de dados está juntamente com o dado) () Não-estruturados (sem esquema de dados, ex.: textos, sites em HTML)
Periodicidade de Produção:	<Ex.: tempo-real, diária, mensal, anual>: Anual
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1 TB): 1,3 MB

Fonte: o autor.

Quadro 20: Fonte de Dados SABESP – Uso Eficiente de Água

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

Cidade:	São Paulo
Domínio	(x) <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador:	DO1-E <Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i>>
Fontes de Dados:	
Fonte 1:	SABESP
Descrição da Fonte:	Dados sobre a taxa de perda de água tratada na cidade de São Paulo. Porcentagem de perda em relação ao total de água produzido na cidade.
Entidade Responsável pelos dados:	ELETRIPAULO Programa Cidades Sustentáveis (Publicador dos dados).

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
Contato	https://www.aeseletropaulo.com.br/Paginas/aes-eletropaulo.aspx
Autor dos dados	--
Colaboradores	--
Data de disponibilização dos dados:	--
Acesso ao dado:	http://www.cidadessustentaveis.org.br
Identificador:	http://indicadores.cidadessustentaveis.org.br/br/SP/sao-paulo/indicadores.xls
Fonte:	--
Idioma:	Português
Relação	<relação específica com outro recursos ou fontes de dados > Identificador do Recurso: Nome do recurso/fonte: URL: Relação: <input type="checkbox"/> IsPartOf HasPart <input type="checkbox"/> IsVersionOf HasVersion <input type="checkbox"/> IsFormatOf HasFormat <input type="checkbox"/> References IsReferencedBy <input type="checkbox"/> IsBasedOn IsBasisFor <input type="checkbox"/> Requires IsRequiredBy
Cobertura Temporal do dado:	2009-2009
Dado bruto ou transformado ?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte> Dado Tratado
Quais softwares são	<ferramentas de software que foram utilizadas para produzir e ou publicar os dados>

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
necessários?	Editor de planilha eletrônica, como LibreOffice Calc, Microsoft Excel.
Dados do domínio da <i>smart city</i>	
Principais fatores atendidos pela fonte:	<i>Sustainable Resource Management/Water e Waste Management</i>
Principais Indicadores e ações atendidos pela fonte:	<i>Efficient use of water use per GDP (Uso eficiente de água)</i>
Características <i>big data</i> e <i>open data</i> (<i>Volume, Variety, Velocity, Variability, Licença Open Data</i> , formato ...):	
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...> Tabular
Natureza dos dados:	() Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) (x) Produzido por Humanos (ex.: <i>posts</i> em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação) () Não Informado
Plataforma da Fonte de dados (meio onde foi coletado):	() Dados coletados de Sensores (x) Web () Outro. Especificar: _____
Detalhamento da plataforma (meio onde foi produzido):	() Serviços Públicos () Dispositivos Móveis (x) Sites () Outro. Especificar: _____
Forma de disponibilização:	() Dados Abertos () <i>Web Services</i> () Postagens em Redes Sociais (x) Outra. Especificar: Planilha do Microsoft Office Excel

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.: coordenadas geográficas) da fonte de dados? () Sim (x) Não
DGA	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito): Secretaria do Meio Ambiente da Prefeitura de São Paulo
Licença Open Data	Possui licença Aberta (Dado Aberto) ? () Sim () Não (x) Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? () Sim () Não (x) Não Informado Especificar: _____
Estrutura do Dado:	(x) Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) () Semiestruturados (ex.: RDF, XML, onde o esquema de dados está juntamente com o dado) () Não-estruturados (sem esquema de dados, ex.: textos, sites em HTML)
Periodicidade de Produção:	<Ex.: tempo-real, diária, mensal, anual>: Anual
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1 TB): 1,3 MB

Fonte: o autor.

Quadro 21: Fonte de Dados Secretaria do Verde e Meio Ambiente - Área Verde

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
Cidade:	São Paulo
Domínio	(x) <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador:	DO1-E <Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i>>

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

Fontes de Dados:

Fonte 1:	SVMA (Secretaria do Verde e Meio Ambiente)
Descrição da Fonte:	Dados sobre a quantidade de área verde por habitante na cidade de São Paulo e sobre
Entidade Responsável pelos dados:	SVMA (Entidade Responsável) Programa Cidades Sustentáveis (Publicador dos dados).
Contato	(11) 3284-2109. http://www.prefeitura.sp.gov.br/cidade/secretarias/meio_ambiente/
Autor dos dados	--
Colaboradores	--
Data de disponibilização dos dados:	--
Acesso ao dado:	http://www.cidadessustentaveis.org.br
Identificador:	http://indicadores.cidadessustentaveis.org.br/br/SP/sao-paulo/indicadores.xls
Fonte:	--
Idioma:	Português
Relação	<relação específica com outro recursos ou fontes de dados > Identificador do Recurso: Nome do recurso/fonte: URL: Relação: () IsPartOf HasPart () IsVersionOf HasVersion () IsFormatOf HasFormat () References IsReferencedBy () IsBasedOn IsBasisFor () Requires IsRequiredBy
Cobertura	2009-2009

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .		
Temporal do dado:		
Dado bruto ou transformado?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte> Dado Tratado	
Quais softwares são necessários?	<ferramentas de software que foram utilizadas para produzir e ou publicar os dados> Editor de planilha eletrônica, como LibreOffice Calc, Microsoft Excel.	
Dados do domínio da <i>smart city</i>		
Principais fatores atendidos pela fonte:	<i>Attractivity of Natural Conditions;</i> <i>Sustainable Resource Management/Water e Waste Management</i>	
Principais Indicadores e ações atendidos pela fonte:	<i>Green space share</i> (Área Verde Por Habitante); <i>Efficient use of water use per GDP</i> (Uso eficiente de água);	
Características <i>big data</i> e <i>open data</i> (<i>Volume, Variety, Velocity, Variability, Licença Open Data</i> , formato ...):		
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...> Tabular	
Natureza dos dados:	<input type="checkbox"/> Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) <input checked="" type="checkbox"/> Produzido por Humanos (ex.: <i>posts</i> em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação) <input type="checkbox"/> Não Informado	
Plataforma da Fonte de dados (meio onde foi coletado):	<input type="checkbox"/> Dados coletados de Sensores <input checked="" type="checkbox"/> Web <input type="checkbox"/> () Outro. Especificar: _____ _____	
Detalhamento da plataforma (meio onde foi)	<input type="checkbox"/> Serviços Públicos <input type="checkbox"/> Dispositivos Móveis <input checked="" type="checkbox"/> Sites	

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

produzido):	(_____) Outro. Especificar: _____ _____
Forma de disponibilização:	() Dados Abertos () <i>Web Services</i> () Postagens em Redes Sociais (x) Outra. Especificar: Planilha do Microsoft Office Excel
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.: coordenadas geográficas) da fonte de dados? () Sim (x) Não
DGA	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito): Secretaria do Meio Ambiente da Prefeitura de São Paulo
Licença <i>Open Data</i>	Possui licença Aberta (Dado Aberto) ? () Sim () Não (x) Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? () Sim () Não (x) Não Informado Especificar: _____ _____
Estrutura do Dado:	(x) Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) () Semiestruturados (ex.: RDF, XML, onde o esquema de dados está juntamente com o dado) () Não-estruturados (sem esquema de dados, ex.: textos, sites em HTML)
Periodicidade de Produção :	<Ex.: tempo-real, diária, mensal, anual>: Anual
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1 TB): 1,3 MB

Fonte: o autor.

Quadro 22: Fonte de Dados Secretaria Municipal de Serviços - Reciclagem

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
Cidade:	São Paulo
Domínio	(x) <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador :	DO1-E <Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i> >
Fontes de Dados:	
Fonte 1:	SES - Secretaria Municipal de Serviços
Descrição da Fonte:	Dados sobre a percentagem de resíduos sólidos reciclados na cidade.
Entidade Responsável pelos dados:	SES (Entidade Responsável) Programa Cidades Sustentáveis (Publicador dos dados).
Contato	(11) 2075-7240 http://www.prefeitura.sp.gov.br/cidade/secretarias/servicos/
Autor dos dados	--
Colaboradores	--
Data de disponibilização dos dados:	--
Acesso ao dado:	http://www.cidadessustentaveis.org.br
Identificador:	http://indicadores.cidadessustentaveis.org.br/br/SP/sao-paulo/indicadores.xls
Fonte:	--
Idioma:	Português
Relação	<relação específica com outro recursos ou fontes de dados> Identificador do Recurso: Nome do recurso/fonte:

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

	URL: Relação: <input type="checkbox"/> IsPartOf HasPart <input type="checkbox"/> IsVersionOf HasVersion <input type="checkbox"/> IsFormatOf HasFormat <input type="checkbox"/> References IsReferencedBy <input type="checkbox"/> IsBasedOn IsBasisFor <input type="checkbox"/> Requires IsRequiredBy
Cobertura Temporal do dado:	2009-2009
Dado bruto ou transformado ?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte> Dado Tratado
Quais softwares são necessários?	<ferramentas de software que foram utilizadas para produzir e ou publicar os dados> Editor de planilha eletrônica, como LibreOffice Calc, Microsoft Excel.
Dados do domínio da <i>smart city</i>	
Principais fatores atendidos pela fonte:	<i>Sustainable Resource Management/Water e Waste Management</i>
Principais Indicadores e ações atendidos pela fonte:	Percentagem de Resíduos Sólidos Reciclados (novo indicador)
Características <i>big data</i> e <i>open data</i> (Volume, Variety, Velocity, Variability, Licença Open Data, formato ...):	
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...> Tabular
Natureza dos dados:	<input type="checkbox"/> Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) <input checked="" type="checkbox"/> Produzido por Humanos (ex.: <i>posts</i> em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação)

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
	() Não Informado
Plataforma da Fonte de dados (meio onde foi coletado):	<input type="checkbox"/> Dados coletados de Sensores <input checked="" type="checkbox"/> Web <input type="checkbox"/> Outro. Especificar: _____ _____
Detalhamento da plataforma (meio onde foi produzido):	<input type="checkbox"/> Serviços Públicos <input type="checkbox"/> Dispositivos Móveis <input checked="" type="checkbox"/> Sites <input type="checkbox"/> Outro. Especificar: _____ _____
Forma de disponibilização:	<input type="checkbox"/> Dados Abertos <input type="checkbox"/> <i>Web Services</i> <input type="checkbox"/> Postagens em Redes Sociais <input checked="" type="checkbox"/> Outra. Especificar: Planilha do Microsoft Office Excel
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.: coordenadas geográficas) da fonte de dados? <input type="checkbox"/> Sim <input checked="" type="checkbox"/> Não
DGA	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito): Secretaria do Meio Ambiente da Prefeitura de São Paulo
Licença <i>Open Data</i>	Possui licença Aberta (Dado Aberto) ? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input checked="" type="checkbox"/> Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input checked="" type="checkbox"/> Não Informado Especificar: _____ _____
Estrutura do Dado:	<input checked="" type="checkbox"/> Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) <input type="checkbox"/> Semiestruturados (ex.: RDF, XML, onde o esquema de dados está juntamente com o dado) <input type="checkbox"/> Não-estruturados (sem esquema de dados, ex.: textos, sites)

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
	em HTML)
Periodicidade de Produção:	<Ex.: tempo-real, diária, mensal, anual>: Anual
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1 TB): 1,3 MB

Fonte: o autor.

Quadro 23: Fonte de Dados Relatório do Ministério do Meio Ambiente – Opinião e Engajamento Público

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
Cidade:	São Paulo
Domínio	(x) <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador:	DO1-E <Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i>>
Fontes de Dados:	
Fonte 1:	Ministério do Meio Ambiente – Governo Federal
Descrição da Fonte:	Dados oriundos uma pesquisa feita pelo Ministério do Meio Ambiente para estudar a opinião pública sobre o meio Ambiente e Consumo Sustentável
Entidade Responsável pelos dados:	Ministério do Meio Ambiente. Secretaria de Articulação Institucional e Cidadania Ambiental.
Contato	Telefone: (061) 2028-1430/1404 http://www.mma.gov.br/destaques/item/8692
Autor dos dados	Samyra Crespo, Eduardo Viola, Fátima Portilho, Pedro Roberto Jacobi, Mauricio Anfré Ribeiro, Hélio Matar, Julia Guivant.
Colaboradores	Marcelo Nascimento. CP2 - Consultoria, Pesquisa e Planejamento.

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
Data de disponibilização dos dados:	Abril 2012
Acesso ao dado:	http://www.mma.gov.br/publicacoes/responsabilidade-socioambiental/category/90-producao-e-consumo-sustentaveis?download=989:o-que-o-brasileiro-pensa-do-meio-ambiente-e-do-consumo-sustentavel
Identificador:	ISBN 978-85-66603-00-2
Fonte:	--
Idioma:	Português
Relação	<relação específica com outro recursos ou fontes de dados > Identificador do Recurso: Nome do recurso/fonte: URL: Relação: <input type="checkbox"/> IsPartOf HasPart <input type="checkbox"/> IsVersionOf HasVersion <input type="checkbox"/> IsFormatOf HasFormat <input type="checkbox"/> References IsReferencedBy <input type="checkbox"/> IsBasedOn IsBasisFor <input type="checkbox"/> Requires IsRequiredBy
Cobertura Temporal do dado:	2012
Dado bruto ou transformado ?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte> Dado tratado, apresentado em relatório e gráficos em PDF .
Quais softwares são necessários?	<ferramentas de software que foram utilizadas para produzir e ou publicar os dados> Leitor de PDF, como Adobe reader.
Dados do domínio da <i>smart city</i>	
Principais fatores atendidos pela fonte:	<i>Environmental Protection</i>

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

Principais Indicadores e ações atendidos pela fonte:	<i>Individual efforts on protecting nature protection</i> (Iniciativas dos cidadãos para preservação do meio ambiente); <i>Opinion on nature protection</i> (opinião sobre a preservação ambiental).
Características <i>big data</i> e <i>open data</i> (<i>Volume, Variety, Velocity, Variability, Licença Open Data</i> , formato ...):	
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...> texto, imagens
Natureza dos dados:	() Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) (x) Produzido por Humanos (ex.: <i>posts</i> em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação) () Não Informado
Plataforma da Fonte de dados (meio onde foi coletado):	() Dados coletados de Sensores (x) Web () Outro. Especificar: _____ _____
Detalhamento da plataforma (meio onde foi produzido):	() Serviços Públicos () Dispositivos Móveis (x) Sites () Outro. Especificar: _____ _____
Forma de disponibilização:	() Dados Abertos () <i>Web Services</i> () Postagens em Redes Sociais (x) Outra. Especificar: Relatório em PDF
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.: coordenadas geográficas) da fonte de dados? () Sim (x) Não
DGA	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito); Ministério do Meio

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
	Ambiente. Secretaria de Articulação Institucional e Cidadania Ambiental
Licença <i>Open Data</i>	Possui licença Aberta (Dado Aberto) ? () Sim () Não (x) Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? () Sim () Não (x) Não Informado Especificar: _____ _____
Estrutura do Dado:	() Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) () Semiestruturados (ex.: RDF, XML, onde o esquema de dados está juntamente com o dado) (x) Não-estruturados (sem esquema de dados, ex.: textos, sites em HTML)
Periodicidade de Produção:	<Ex.: tempo-real, diária, mensal, anual>: Aproximadamente de 5 em 5 anos.
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1 TB): 6 MB

Fonte: o autor.

Quadro 24: Fonte de Dados Redes Sociais – Amostra de Postagens

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
Cidade:	São Paulo
Domínio	(x) <i>Environment</i> () <i>Mobility</i> () <i>Government</i> () <i>Economy</i> () <i>People</i> () <i>Living</i>
Identificador:	DO1-E <Ex.: DO2-M, se o domínio escolhido for <i>Mobility</i> >
Fontes de Dados:	
Fonte 1:	Twitter
Descrição da	Dados extraídos da base de postagens na rede social Twitter

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

Fonte:	envolvendo as palavras-chaves “proteção ambiental”, postados na cidade de São Paulo. Obs.: quando utilizado o filtro por cidade a base retorna um número extremamente baixo.
Entidade Responsável pelos dados:	Twitter
Contato	https://twitter.com
Autor dos dados	--
Colaboradores	--
Data de disponibilização dos dados:	--
Acesso ao dado:	https://twitter.com/search?q=proteção%20ambiental%20lang%3Apt%20near%3A\"São%20Paulo\"%20within%3A15mi%20since%3A2009-02-03%20until%3A2016-01-12&src=typd&lang=pt
Identificador :	https://twitter.com/search?q=proteção%20ambiental%20lang%3Apt%20near%3A\"São%20Paulo\"%20within%3A15mi%20since%3A2009-02-03%20until%3A2016-01-12&src=typd&lang=pt
Fonte:	--
Idioma:	Português
Relação	<relação específica com outro recursos ou fontes de dados> Identificador do Recurso: Nome do recurso/fonte: URL: Relação: () IsPartOf HasPart () IsVersionOf HasVersion

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i> .	
	<input type="checkbox"/> IsFormatOf HasFormat <input type="checkbox"/> References IsReferencedBy <input type="checkbox"/> IsBasedOn IsBasisFor <input type="checkbox"/> Requires IsRequiredBy
Cobertura Temporal do dado:	2009-2016
Dado bruto ou transformado ?	<o dado sofreu processo de tratamento (ETL) ou foi coletado diretamente na fonte> Dado bruto, coletado no site Twitter
Quais softwares são necessários?	<ferramentas de software que foram utilizadas para produzir e ou publicar os dados> Browser, como o Chrome.
Dados do domínio da <i>smart city</i>	
Principais fatores atendidos pela fonte:	<i>Environmental Protection</i>
Principais Indicadores e ações atendidos pela fonte:	<i>Opinion on nature protection</i> (opinião sobre a preservação ambiental).
Características <i>big data</i> e <i>open data</i> (<i>Volume, Variety, Velocity, Variability, Licença Open Data</i> , formato ...):	
Tipo do dado:	<categoria do recurso, como texto, imagem, som, dados exportados em listas ou tabelas ...> Texto
Natureza dos dados:	<input type="checkbox"/> Produzido por Máquinas (ex.: dados de sensores, Internet das Coisas) <input checked="" type="checkbox"/> Produzido por Humanos (ex.: posts em redes sociais, textos extraídos de sites, dados informados por usuários de sistemas de informação) <input type="checkbox"/> Não Informado
Plataforma da Fonte de dados (meio onde foi)	<input type="checkbox"/> Dados coletados de Sensores <input checked="" type="checkbox"/> Web <input type="checkbox"/> Outro.

Formulário de Mapeamento de Proveniência de Dados para *smart cities* baseadas em *big data* e *open data*.

coletado):	Especificar: _____ _____
Detalhamento da plataforma (meio onde foi produzido):	() Serviços Públicos () Dispositivos Móveis (x) Sites () Outro. Especificar: _____ _____
Forma de disponibilização:	() Dados Abertos () <i>Web Services</i> (x) Postagens em Redes Sociais () Outra. Especificar: Relatório em PDF
Dados Contextuais (localização do dado)	Possui informações sobre a localização (ex.:coordenadas geográficas) da fonte de dados? (x) Sim () Não
DGA	Caso for Dado Governamental, descrever de qual agência (ex.: Departamento Municipal de Trânsito):
Licença <i>Open Data</i>	Possui licença Aberta (Dado Aberto) ? () Sim () Não (x) Não Informado Caso não for Dado Aberto, possui algum tipo de restrição legal de uso? () Sim () Não (x) Não Informado Especificar: _____ _____
Estrutura do Dado:	() Dados Estruturados (ex.: dados extraídos de sistemas de informação, com esquema de dados separado do dado) () Semiestruturados (ex.: RDF, XML, onde o esquema de dados está juntamente com o dado) (x) Não-estruturados (sem esquema de dados, ex.: textos, sites em HTML)
Periodicidade de Produção:	<Ex.: tempo-real, diária, mensal, anual>: Diária
Tamanho	Tamanho Aproximado da amostra coletada (ex.: 100 MB, 1 GB, 1

Formulário de Mapeamento de Proveniência de Dados para <i>smart cities</i> baseadas em <i>big data</i> e <i>open data</i>.	
ho	TB): 144 KB

Fonte: o autor.

Abaixo, seguem as principais informações encontradas nas fontes de dados mapeadas.

Quadro 25: Informações Obtidas para Análise

Fator <i>Attractivity of Natural Conditions</i>		
Indicador:	Fonte:	Valor:
<i>Green space share</i> (Área Verde Por Habitante);	SVMA (Secretaria do Verde e Meio Ambiente); Dados em planilhas.	12 m ² / habitante
Fator: <i>Sustainable Resource Management/Water e Waste Management</i>		
Indicador:	Fonte:	Valor:
<i>Efficient use of water use per GDP</i> (Uso eficiente de água)	SABESP Dados em planilhas.	34,94 %
Indicador:	Fonte:	Valor:
Reciclagem de Resíduos Sólidos	SES (Secretaria Municipal de Serviços) Dados em planilhas.	1,02%
Indicador:	Fonte:	Valor:
<i>Efficient use of electricity use per GDP</i> (Uso eficiente de eletricidade)	ELETROPAULO Dados em planilhas.	245,83 kw/h/mês
Fator <i>Environmental Protection</i>		
Indicador:	Fonte:	Valor:
<i>Individual efforts on protecting nature protection</i> (Iniciativas dos cidadãos para	Ministério do Meio Ambiente Relatório em PDF	80% da população não participa de atividades em prol do meio ambiente

preservação do meio ambiente)		
<i>Opinion on nature protection</i> (opinião sobre a preservação ambiental).	Ministério do Meio Ambiente Relatório em PDF	Por que cuidar do meio ambiente? Sobrevivência - 65% população Futuro Melhor – 15 % Prevenção de Catástrofes – 4%
<i>Posts em Redes Sociais</i>	Postagens no Twitter sobre o assunto proteção ambiental. Filtro: cidade a palavra-chave: Proteção ambiental	8 postagens para a cidade de São Paulo (número pequeno devido ao fato de que poucos usuários informam sua localização na rede social utilizada)

Fonte: o autor.

Com estas informações como apoio, um gestor de uma *smart city* possui maiores subsídios para os processos de tomada de decisões, realizar processos de *benchmarking* entre cidades, avaliar a performance de sua cidade em diversas áreas importantes para seu desenvolvimento, dentre outras tarefas. O uso dos artefatos neste experimento auxiliou principalmente a registrar a proveniência das fontes de dados, buscar por novas fontes de dados e a construir novos indicadores, baseados nos dados encontrados. A seguir, são apresentadas as conclusões deste trabalho.

5 CONCLUSÕES

O problema de pesquisa inicialmente proposto questiona como gerar conhecimento para apoiar as *smart cities*, utilizando *big data* e *open data*. Preocupou-se também nesta pesquisa com os seguintes questionamentos: quais dados devem ser coletados, quais informações devem ser produzidas, quais fontes de dados utilizar e quais são os principais atores neste processo. Para solucionar estes problemas, foi desenvolvido o Modelo Conceitual proposto, que mapeou as principais fontes de dados, suas características e atributos de proveniência, e as informações a serem extraídas, através da indicação de conjuntos de metadados que guiam estas tarefas. Sendo assim, administradores de cidades, que visem iniciar projetos de *smart cities*, ou sistemas de informação para estas cidades por exemplo, podem se beneficiar do Modelo proposto antes de iniciar seus projetos.

Como contribuições acadêmicas, nos procedimentos metodológicos desta pesquisa, mais precisamente na revisão bibliográfica, foi realizada uma revisão teórica sobre os temas envolvidos: *smart cities*, *big data*, *open data*, governança de dados, qualidade de dados e informação e modelos conceituais.

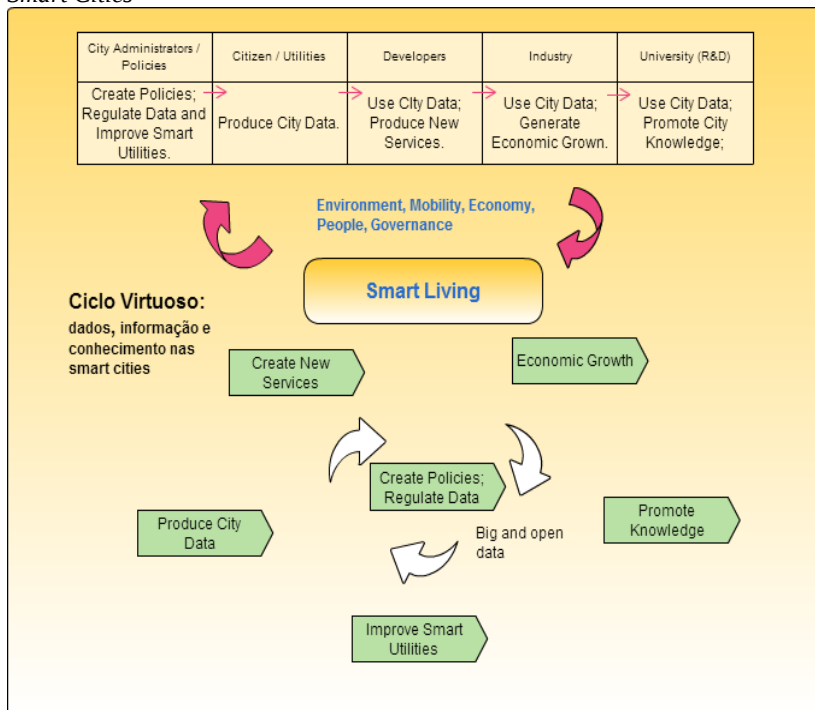
Em relação às *smart cities*, foi identificado que este conceito promove o uso de TICs como meio de melhorar a qualidade de vida nos problemáticos centros urbanos atuais. Porém, tal abordagem também sugere que é necessário investir em capital humano e social, transparência governamental, economia criativa, dentre outros fatores. Estes fatores são comumente agrupados em seis domínios de atuação: *Environment* (Recursos Naturais), *Mobility & ICT* (Mobilidade Urbana e TICs), *Economy* (Competitividade Econômica), *Government* (Transparência), *People* (Capital Humano e Social) e *Living* (Qualidade de Vida), sendo que estes domínios são classificados na literatura em dois grupos: *Hard* e *Soft*, conforme varia a intensidade do impacto das tecnologias baseadas em *big data* nestes grupos.

Foi visto também que as *smart cities* produzem e consomem grandes volumes de dados e informações. Em outras palavras, elas estão estreitamente relacionadas com o fenômeno conhecido como *big data*. Na pesquisa exploratória realizada foi identificado que este fenômeno possui três características principais: seu grande volume de dados, a velocidade com que este volume aumenta, e a variedade de formatos utilizados pelas diversas fontes (*volume*, *velocity* e *variety*,

respectivamente), e que alguns autores corroboram este conceito definindo que analistas de *big data* precisam se preocupar com a veracidade dos dados a serem utilizados (*veracity*), com a variabilidade dos esquemas de dados e suas interpretações (*variability*), e o valor das informações (*value*) contidas nestes dados. Para auxiliar nestas dimensões, podem ser aplicados conceitos de qualidade de dados e governança de dados, mapeando a proveniência das fontes de dados por exemplo. Além da caracterização deste fenômeno, foi identificado na literatura que suas causas principais são o aumento do uso de TICs em diversos setores da sociedade, TICs estas com cada vez maiores capacidades de armazenamento e processamento de dados, o que consequentemente aumenta a disponibilidade de dados e informações a serem utilizadas.

Dentro das fontes de dados do fenômeno *big data*, focou-se e levantou-se as características das fontes *Open Data* (dados abertos). Foi apontado que o uso de dados abertos potencializa a utilização do fenômeno *big data*, pois seus dados passam a ter legitimidade legal para sua obtenção e uso, além de apresentarem uma maior interoperabilidade, o que facilita o processo de extração de informação e conhecimento com base nestas fontes. Foi identificado também que o movimento *Open Data* traz diversos benefícios para a sociedade e é estratégico para as *smart cities*, pois incentiva a transparência, controle social, participação popular, geração de novos serviços, geração de conhecimento e inovação (OPEN DATAHANDBOOK), e que tais vantagens são diretamente relacionadas aos apontamentos das *smart cities*, pois estão ligadas diretamente aos domínios *Governance* (transparência), *People* (participação na vida pública, geração de conhecimento) e *Economy* (empreendedorismo).

Figura 13: Ciclo Virtuoso de Uso e Produção de Dados e Conhecimento nas *Smart Cities*



Fonte: o autor.

Ainda em relação às contribuições acadêmicas da pesquisa exploratória desta dissertação, foi identificado que, ao utilizar fontes *big data* e *open data*, os administradores das *smart cities* criam um ciclo virtuoso de criação, utilização e compartilhamento de dados e conhecimento sobre estas cidades (Figura 13).

Este ciclo é iniciado pelos administradores das cidades (*City Administrators*), que criam as políticas públicas (*Policies*) e regulam o uso dos dados (*Regulate data*). Estes dados auxiliam a melhorar os serviços públicos, através de sua análise e melhor eficiência com uso de TICs, e servem como base para geração de novos serviços de informação, bem como auxiliam a promover conhecimento sobre o funcionamento e comportamento destas cidades, uma vez que os dados produzidos contém informações oriundas principalmente dos serviços públicos destas cidades, suas agências governamentais, sobre seus

habitantes e seu meio.

Neste contexto, a contribuição maior do Modelo proposto foi em auxiliar o mapeamento de fontes de dados *big data* e *open data* nos seis domínios identificados nas *smart cities*, sugerindo metadados para isso. Através de seu uso, o Modelo retorna dois conjuntos de metadados. O primeiro conjunto identifica as informações a serem analisadas em cada domínio das *smart cities*. O segundo conjunto faz o mapeamento das principais fontes de dados, registrando sua proveniência e características baseadas no fenômeno *big data* e movimento *open data*, e indicando onde procurar por estas fontes. Para utilizar o Modelo, foram desenvolvidos sete formulários de entrada de dados, onde o seu utilizador pode informar qual domínio ele deseja analisar, e então, receber os conjuntos de metadados que o auxiliam nesta tarefa. Com esta resposta em mãos, é possível realizar a construção de indicadores para a análise do funcionamento destas cidades, identificar ações comumente necessárias em *smart cities*, ou ainda identificar lacunas de bases de dados a serem desenvolvidas pelos administradores de cidades.

Em relação às dificuldades encontradas, a questão de trabalhar com uma área relativamente nova, tendo autores de 2014 e 2015 sendo citados como referência por exemplo, forçou este pesquisador a estar sempre em contato com publicações recentes que complementassem sua fundamentação teórica e, conseqüentemente, modelo proposto.

Como trabalhos futuros, propõe-se o desenvolvimento de um guia de utilização dos artefatos deste Modelo. Estes também podem ser implementados em formato de um sistema de informação, para automatizar a sua utilização, retornando em interfaces gráficas o resultado trazido pelo Modelo. Sugere-se também adição de novos atributos dos conjuntos de metadados de cada um dos domínios das *smart cities*, a ser feito com especialistas de cada domínio (ex.: sociólogos, engenheiros ambientais etc), para somar às informações trazidas pelo Modelo, bem como a identificação, em forma de modelagem, das relações entre os atributos dos seis domínios (ex.: meio ambiente, transportes, e taxas de emissão de carbono).

Sugere-se também aplicar este Modelo como um guia em projetos de diversas naturezas, como por exemplo, projetos de representação de conhecimento através do desenvolvimento de ontologias de domínio, estudos de caso de iniciativas de *smart cities* em determinadas cidades, criação de bases de dados, sistemas de informação e construção de indicadores baseados em *big data* e *open data* para as *smart cities*, dentre outras aplicações a serem exploradas por futuros pesquisadores.

REFERÊNCIAS

AGUNE, Roberto Meizi; GREGORIO FILHO, Alvaro Santos; BOLLIGER, Sergio Pinto. GOVERNO ABERTO SP: DISPONIBILIZAÇÃO DE BASES DE DADOS E INFORMAÇÕES EM FORMATO ABERTO. In: **CONGRESSO CONSAD DE GESTÃO PÚBLICA**, 3., 2010, Brasília. Painel. Brasília: Congresso Consad de Gestão Pública, 2010. p. 1 - 36. Disponível em: <[http://www.prefeitura.sp.gov.br/cidade/secretarias/upload/controladoria_geral/arquivos/C3_TP_GOVERNO ABERTO SP DISPONIBILIZACAO DE BASES DE DADOS.pdf](http://www.prefeitura.sp.gov.br/cidade/secretarias/upload/controladoria_geral/arquivos/C3_TP_GOVERNO_ABERTO_SP_DISPONIBILIZACAO_DE_BASES_DE_DADOS.pdf)>. Acesso em: 05 jul. 2015.

ALBINO, Vito; BERARDI, Umberto; DANGELICO, Rosa Maria. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. **Journal Of Urban Technology**, [s. l.], v. 22, n. 1, p.3-21, 2 jan. 2015. Informa UK Limited. DOI: 10.1080/10630732.2014.942092.

ALLIANCE, University. **What is Big Data?** Disponível em: <http://www.villanovau.com/resources/bi/what-is-big-data/#.VBZnxZ_7G01>. Acesso em: 20 ago. 2014.

ALONSO, José M.. **Introducing the Open Data Index**. 2012. Disponível em: <<http://webfoundation.org/2012/09/introducing-the-open-data-index/>>. Acesso em: 04 ago. 2015.

ASSIS, Juliana de; MOURA, Maria Aparecida. A Qualidade da Informação na Web: uma abordagem semiótica; La Calidad de la Información en la Web. **Informação & Informação**, Londrina, v. 16, n. 3, p.96-117, 8 maio 2011. Universidade Estadual de Londrina. DOI: 10.5433/1981-8920.2011v16n3p96.

BARATA, André Montóia. **Governança de dados em organizações brasileiras: uma avaliação corporativa entre os benefícios previstos na literatura e os obtidos pelas organizações**. 2015. 154 f. Dissertação (Mestrado) - Curso de Pós Graduação em Sistemas de Informação, Universidade de São Paulo, São Paulo, 2015.

BARBIERI, Carlos. **Uma visão sintética e comentada do Data**

Management Body of Knowledge (DMBOK). Belo Horizonte: Fumsoft, 2013. 45 p.

BATINI, C.; CERI, S.; e NAVATHE, S. B. Conceptual database design: an entity-relationship approach. Redwood: Benjamin/Cummings, 1992.

BEYER, Mark A.; LANEY, Douglas. **The Importance of 'Big Data': A Definition.** 2012. Disponível em: <<https://www.gartner.com/doc/2057415/importance-big-data-definition>>. Acesso em: 27 set. 2014.

BERNERS-LEE, Tim. **Linked Data: Design Issues** 2006. Disponível em <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 20 jun. 2015.

BENTANCOURT, Silvia Maria Puentes. **SERVQUAL COMO INSTRUMENTO DE GESTÃO DA QUALIDADE DA INFORMAÇÃO EM AMBIENTE EAD.** 2015. 108 f. Dissertação (Mestrado) - Curso de Pós-graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2015.

BETTINO, Larry A., **Transforming Big Data Challenges Into Opportunities.** Information Management. Extraído de<<http://bit.ly/PWKQtq><http://bit.ly/PWKQtq>>. Acessado em 18/08/2014.

BIVAR, Bárbara et al., **Uma Comparação entre os Modelos de Proveniência OPM e PROV.** BDBComp. Disponível em <<http://www.lbd.dcc.ufmg.br/colecoes/escience/2013/002.pdf>>. Acessado em: 18/01/2016.

BOTELHO, Louise de Lira Roedel. **Aprendizagem Gerencial na Mudança em uma Organização Intensiva em Conhecimento.** 2012. 262 f. Tese (Doutorado) - Curso de Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Departamento de Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2012. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2013/01/LOUISEBOTELHO.pdf>>. Acesso em: 14 ago. 2014.

BRASIL. **Lei no 12.527**, de 18 de novembro de 2011. In: Diário Oficial da República Federativa do Brasil, Brasília, DF, 18 nov. 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em 14/09/2014.

BRASIL. Lei nº 12.527, de 18 de janeiro de 2011. **Lei de Acesso à Informação**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Acesso em: 10 ago. 2015.

BRASIL. GOVERNO FEDERAL, Governo. **Portal Brasileiro de Dados Abertos**. 2015. Disponível em: <<http://dados.gov.br/aplicativos/>>. Acesso em: 15 jul. 2015.

BRASIL. GOVERNO FEDERAL, Governo. O que são dados abertos? **Portal Brasileiro de Dados Abertos**. 2015. Disponível em: <<http://dados.gov.br/dados-abertos/>>. Acesso em: 17 ago. 2015.

BRANCHEAU, James C.; WETHERBE, James C.. INFORMATION ARCHITECTURES: METHODS AND PRACTICE. **Information Processing And Management**. Minesota, p. 453-463. maio 1986.

BRITTO, Mozart da Silva; ALMEIDA JÚNIOR, Jorge Rady de. QUALIDADE DE DADOS PARA DATA WAREHOUSE: ROTEIRO DE IMPLEMENTAÇÃO. In: CONGRESSO INTERNACIONAL DE GESTÃO DA TECNOLOGIA E SISTEMAS DE INFORMAÇÃO, 3., 2006, São Paulo. **Anais...** . São Paulo: Tecsi - Laboratório de Tecnologia e Sistemas de Informação Fea Usp, 2006. p. 255 - 272.

BUNEMAN, Peter; KHANNA, Sanjeev; WANG-CHIEW, Tan. Why and Where: A Characterization of Data Provenance. **Database Theory — Icdt 2001**, [s.l.], v. 5, n. 8, p.316-330, nov. 2001. Springer Science + Business Media. DOI: 10.1007/3-540-44503-x_20.

BUNGE, M. **Treatise on basic philosophy**. Part. II. Boston: D. Reidel, 1985, v. 7.

BUNGE, Mario. **Emergence and Convergence**: Qualitative Novelty and the Unity of Knowledge. Toronto: National Library Of Canada, 2003.

CARAGLIU, Andrea; BO, Chiara del; NIJKAMP, Peter. Smart Cities in Europe. **Journal Of Urban Technology**. Londres, p. 65-82. 10 ago. 2011. Disponível em: <http://www.tandfonline.com/doi/abs/10.1080/.VBZNMx_Hk8o#.VBZPMZ_7G00>. Acesso em: 11 jun. 2014.

CARVALHO, Isamir Machado de. **A DINÂMICA DOS MECANISMOS DE PROTEÇÃO E COMPARTILHAMENTO DE CONHECIMENTO, NO PROCESSO DE DESENVOLVIMENTO DE SOFTWARE, EM UMA EMPRESA PÚBLICA DE TECNOLOGIA DA INFORMAÇÃO (TI)**. 2014. 288 f. Tese (Doutorado) - Curso de Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2014. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2015/04/Isamir-Machado-de-Carvalho.pdf>>. Acesso em: 13 jul. 2015.

CHAHIN, A. et al. E-Gov.br: a próxima revolução brasileira: eficiência, qualidade e democracia: o governo eletrônico no Brasil e no mundo. **Prentice Hall**. São Paulo, 2004.

CHEN, Min; MAO, Shiwen; LIU, Yunhao. Big Data: A Survey. **Springer**, Estados Unidos, v. 2,n.19,p.171-209,Abr.2014. Disponível em: <<http://link.springer.com/article/10.1007%2Fs11036-013-0489-0>>. Acesso em: 13 set. 2014.

COHILL, A. Information architecture and the design process Taking software design seriously. **Academic Press Professional**, Inc. San Diego, CA, USA. 1991. ISBN:0-12-397710-X.

COSTA, Luciano Antonio. **SISTEMA DE CONHECIMENTO PARA APOIO E GESTÃO DE RECURSOS DE PROGRAMAS DE GOVERNO EM MUNICÍPIOS**. 2013. 127 f. Dissertação (Mestrado) - Curso de Engenharia e Gestão do Conhecimento, Ppgec, Ufsc, Florianópolis, 2013. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2014/01/Luciano-Antonio-Costa.pdf>>. Acesso em: 22 set. 2015.

CUPANI, Alberto. La peculiaridad del conocimiento tecnológico. **Scientiae Studia**, São Paulo, v. 4, n. 3, p. 353-71, 2006.

DAMA. **Body of Knowledge**. 2015. Disponível em:

<<https://www.dama.org/content/body-knowledge>>. Acesso em: 26 dez. 2015.

DAVENPORT, Thomas. **Big Data at Work: Dispelling the Myths, Uncovering the Opportunities**. Boston: Harvard Business Publishing Corporation, 2014. 229 p.

DAVENPORT, Thomas; PRUSAK, Laurence. Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual. **Editora Campus**, 1998.

DINIZ, Vagner. COMO CONSEGUIR DADOS GOVERNAMENTAIS ABERTOS. In: III CONGRESSO CONSAD DE GESTÃO PÚBLICA, 3., 2010, Brasília - Df. Painel. **III Congresso Consad de Gestão Pública. Brasília - DF: Consad**, 2010. Disponível em: <https://i3gov.planejamento.gov.br/como_conseguir_dados_governamentais_abertos.pdf>. Acesso em: 17 ago. 2015.

DOBRE, Ciprian; XHAFA, Fatos. Intelligent services for Big Data science. **Future Generation Computer Systems**. Amsterdam, p. 267-281. jul. 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167739X13001593>>. Acesso em: 05 set. 2014.

DUTTON, W., Blumler, J., KRAEMER, K. (1987) (eds.) **Wired Cities: Shaping the Future of Communications**. Nova York: G.K. Hall.

EDWARDS, Justin. **Web Foundation Launches the Web Index**. 2012. Disponível em: <<http://webfoundation.org/2012/09/web-foundation-launches-the-web-index/>>. Acesso em: 04 ago. 2015.

EGC. **Linhas de Pesquisa**. Disponível em: <<http://www.egc.ufsc.br/pesquisas/linhas-de-pesquisa/>>. Acesso em: 13 nov. 2015.

EOEKE. **Plataforma para gerenciamento de indicadores**. Disponível em: <<http://eokoe.github.io/Iota/>>. Acesso em: 14 out. 2015.

ETZKOWITZ, H.; LEYDESDORFF, L. The dynamics of innovation: from national systems and ‘Mode 2’ to a triple helix of universityindustry-government relations. **Research Policy**, vol. 29, p.

109-23, 2000.

ESTADOS UNIDOS. WHITE HOUSE. **Federal Enterprise Architecture**. Disponível em: <<https://www.whitehouse.gov/omb/e-gov/fea>>. Acesso em: 09 jul. 2015.

FALCONER, Gordon; MITCHELL, Shane. Smart City Framework: A Systematic Process for Enabling Smart Connected Communities. **Cisco Internet Business Solutions Group (ibsg)**. California, p. 9-12. set. 2012.

FAN, Wei; BIFET, Albert. Mining Big Data: Current Status, and Forecast to the Future. **SIGKDD Explorations**, China, v. 2, n. 14, p.1-5, mar. 2012. Disponível em: <<http://www.kdd.org/sites/default/files/issues/14-2-2012-12/V14-02-01-Fan.pdf>>. Acesso em: 15 set. 2014.

FENSEL, D.; HERMELEN, F. van. **On-To-Knowledge: Content-Driven Knowledge Management Tools through Evolving Ontologies**

FERNANDES, Agnaldo Aragon; ABREU, Vladimir Ferraz de. **Implantando a Governança de TI: da estratégia à gestão de processos e serviços**. 3. ed. Rio de Janeiro: Brasport, 2012.

FLORIANÓPOLIS. PREFEITURA DE FLORIANÓPOLIS. **DEMONSTRATIVO DA EXECUÇÃO DAS DESPESAS POR FUNÇÃO/SUBFUNÇÃO: JANEIRO A OUTUBRO 2012/ BIMESTRE SETEMBRO-OUTUBRO**. 2012. Disponível em: <<http://www.pmf.sc.gov.br/entidades/planejamento/index.php?pagina=govgestao&menu=2&tprel=359&pg=2>>. Acesso em: 28 jan. 2014.

FREIXO, Jorge; ROCHA, Álvaro. Arquitetura de Informação de Suporte à Gestão da Qualidade em Unidades Hospitalares. **Iberian Journal Of Information Systems And Technologies**, Coimbra, Portugal, v. 12, n. 14, p.1-16, 10 nov. 2014. Anual. AISTI. DOI: 10.17013/risti.14.1-15.

FREITAS JUNIOR, Vanderlei et al. A pesquisa científica e tecnológica. **Espacios**, [S.I.], v. 35, n. 9, p.12-12, jul. 2014. Disponível em: <<http://www.revistaespacios.com/a14v35n09/14350913.html>>. Acesso

em: 16 out. 2014.

GANTZ, John; REINSEL, David. Extracting Value from Chaos. **IDC Iview**, Framingham, v. 01, n. 17, p.1-12, jun. 2011. Disponível em: <<http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>>. Acesso em: 13 set. 2014.

GARTNER. **Big Data**. IT Glossary [20--]. Disponível em <<http://www.gartner.com/it-glossary/big-data/>>. Acessado em 19/08/2014.

GIFFINGER, Rudolf et al. **Smart cities Ranking of European medium-sized cities**. Vienna: Vienna University Of Technology, 2007. 28 p. Disponível em: <http://www.smart-cities.eu/download/smart_cities_final_report.pdf>. Acesso em: 06 jun. 2014.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4a. ed. São Paulo: Atlas, 2002.

GÓMEZ-PÉREZ, A; et al. Ontologic Engineering: with examples from the areas of knowledge management, e-commerce and the semantic web. **Springer-Verlag**, 2004.

GOVERNOELETRÔNICO, Portal. **O que é E-GOV**. 2005. Disponível em: <http://www.governoeletronico.com.br/index.php?option=com_content&task=view&id=19&Itemid=29>. Acesso em: 17 ago. 2015.

GOVERNMENT, U.S.. **U.S. Government's Open Data**. 2015. Disponível em: <<http://www.data.gov/>>. Acesso em: 15 jul. 2015.

GROBELNIK, Marko. **Big Data Tutorial**. Kalamaki: Jožef Stefan Institute, 2012. Color. Disponível em: <http://videolectures.net/eswc2012_grobelnik_big_data/>. Acesso em: 22 jul. 2014.

IBGE, Instituto Brasileiro de Geografia e Estatística. IBGE divulga as estimativas populacionais dos municípios em 2014. 2014. Disponível em: <<http://saladeimprensa.ibge.gov.br/noticias?view=noticia&id=1&busca=1&idnoticia=2704>>. Acesso em: 27 jan.

2015.

IBM, Software. **The IBM big data platform**. Somers: IBM Software, 2013. 4 p. Disponível em: <<http://public.dhe.ibm.com/common/ssi/ecm/en/imb14135usen/IMB14135USEN.PDF>>. Acesso em: 15 ago. 2014.

IBM. Bussiness System Planning - **Informaton Systems Planning Guide**. 1984.IBM products.

IBM.The Four Vs of Big Data.**IBM Big Data & Analytcs Hub**.Disponível em <<http://www.ibmbigdatahub.com/infographic/four-vs-big-data>>. Acesso em 22 set. 2015.

IBM. **The IBM Data Governance Council Maturity Model: Building a roadmap for effective data governance**. Nova York: IBM, 2007. 16 p.

INTELLIGENT COMMUNITY FORUM – ICF (New York). **Intelligent Community Indicators**. Disponível em: <https://www.intelligentcommunity.org/index.php?src=gendocs&ref=Research_Intelligent_Community_Indicators&category=Research>. Acesso em: 28 maio 2015.

KERN, Vinicius Medina. O SISTEMISMO DE BUNGE: FUNDAMENTOS, ABORDAGEM METODOLÓGICA E APLICAÇÃO A SISTEMAS DE INFORMAÇÃO. In: **XII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, Comunicação Oral. Distrito Federal: Enancib, 2011.

KESMODEL, David. **Monsanto to Buy Climate Corp. for \$930 Million**. 2013. Disponível em: <<http://www.wsj.com/articles/SB10001424052702304176904579111042198455468>>. Acesso em: 05 ago. 2015.

KETTINER, William; TENG, James; GUHA, Subashish. Information architectural design in business process reengineering. **Journal Of Information Technology. United Kingdom Of Great Britain & Northern Ireland**, p. 27-37. mar. 1996.

KUK, George; JANSSEN, Marijn. The Business Models and Information Architectures of Smart Cities. **Journal Of Urban Technology**. United Kingdom, p. 39-52. 02 abr. 2011

LABORATÓRIO BRASILEIRO DE CULTURA DIGITAL (Org.). **Manual dos dados abertos**: governo. S.i: W3c, 2011. 58 p. Disponível em: http://www.w3c.br/pub/Materiais/PublicacoesW3C/Manual_Dados_Abertos_WEB.pdf>. Acesso em: 07 ago. 2015.

LATHROP, Daniel; RUMA, Laurel. Open Government: Transparency, Collaboration and Participation in Practice. **O'Reilly Media**. 2010. ISBN 978-0-596-80435-0.

LETAIFA, S. Ben. How to strategize smart cities: Revealing the SMART model. **Journal of Business Research**. Elsevier. Université du Québec à Montréal, Canada. 2015.

LONDEIX, B. Evaluating the quality of entity relationship models, Information and Software Technology, V.37, 1995.

LUCHESI, Eunice. GESTÃO DO CONHECIMENTO NAS ORGANIZAÇÕES. 2012. **Companhia de Engenharia de Tráfego de São Paulo**. NT 221. São Paulo.

LODKEM - LINKED OPEN DATA & KNOWLEDGE ENGINEERING AND MANAGEMENT. **Aplicações**. Disponível em: <http://lodkem.egc.ufsc.br/index.php/home/aplicacoes>>. Acesso em: 15 jul. 2015.

LOMBARDIA, Patrizia et al. Modelling the smart city performance. **Innovation: The European Journal of Social Science Research**. [s. l.], p. 137-149. jun. 2012.

MANYIKA, James et al. **Big Data: The next frontier for innovation, competition, and productivity**. Nova York: McKinsey Global Institute, 2011. 20 p. Disponível em: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>. Acesso em: 18 jun. 2014.

MARINS, André. **Modelos Conceituais para Proveniência**. Dissertação de Mestrado. Programa de PósGraduação em Informática da PUC-Rio. 2008.

MARTIN, Andrew; DMITRIEV, Dmitry; AKEROYD, John. A resurgence of interest in Information Architecture. **International Journal Of Information Management**, [s.l.], v. 30, n. 1, p.6-12, fev. 2010. Elsevier BV. DOI: 10.1016/j.ijinfomgt.2009.11.008. Disponível em: <<http://api.elsevier.com/content/article/PII:S0268401209001455?httpAccept=text/xml>>. Acesso em: 18 ago. 2015.

MELLO, Ronaldo dos Santos, et al. Dados semi-estruturados. **XV SIMPÓSIO BRASILEIRO DE BANCO DE DADOS** (2000). Tutorial. Disponível em <ftp://ftp.cse.buffalo.edu/users/azhang/disc/disc01/cd1/out/websites/sbbd/tutoriais_sbbd.htm#Tutorial3>. Acesso em: 15 ago. 2015.

MATHEUS, Ricardo; VAZ, José Carlos; RIBEIRO, Manuella Maia. Dados Governamentais Abertos e o ciclo virtuoso de uso de dados e informações públicas para o controle social, melhoria dos serviços públicos e desenvolvimento da economia: Estudo de Caso da Cidade do Rio de Janeiro. In: **WORKSHOP DE TRANSPARÊNCIA EM SISTEMAS**, 1., 2014, Londrina. **Anais do Workshop de Transparência em Sistemas**. Londrina: Puc, 2014. Disponível em: <http://wtrans.inf.puc-rio.br/WTRANSartigos/pdf_counter.lua?wtrans=WTRANS14&file_name=paper8.pdf>. Acesso em: 10 ago. 2015.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. **Big Data: A Revolution that Will Transform how We Live, Work, and Think**. Nova York: Houghton Mifflin Harcourt Publishing Company, 2013. 231 p.

MORVILLE, Peter; ROSENFELD, Louis. **Information Architecture: for the World Wide Web**. 3. ed. Paris: O'reilly, 2007. 528 p. Disponível em: <<http://yunus.hacettepe.edu.tr/~tonta/courses/fall2010/bby607/IAWW.pdf>>. Acesso em: 06 jul. 2015.

MOSLEY, M. et al. **The DAMA Guide to The Data Management Body of Knowledge: DAMA - DMBOK Guide**. 1. ed. Estados Unidos: Technics Publications, 2009.

NAM Taewoo; PARDO Theresa A. Conceptualizing Smart City with Dimensions of Technology, People, and Institutions. In: ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH, 12th, 2011, New York. **Proceedings**. New York: Acm Magazines And Online Publications, 2011. 282-291.

NAVATHE, S. B. Evolution of Data Modeling for Databases. Communications of the ACM, 35(9):112-123, 1992.

NEIROTTI, Paolo et al. Current trends in Smart City initiatives: Some stylised facts. **Cities: the international journal of urban policy and planning**. Amsterdã, p. 25-36. jun. 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0264275113001935>>. Acesso em: 15 set. 2014.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka. Criação do conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação. 2. Ed. **Rio de Janeiro**: Campus, 1997.

OLIVEIRA, Álvaro de. **Human Smart Cities in Europe**. Alameda - Portugal: Alfamicro, 2013. 60 slides,color. Disponível em: <http://www.egov.ufsc.br/portal/sites/default/files/human_smart_cities_in_europe-_junho_alvaro_2013-brasil_final.pdf>. Acesso em: 14 set. 2015.

O'REILLY, Media. **Big Data Now**. Sebastopol: O'reilly Media, 2012. 119 p.

OPEN DATA HANDBOOK. **Por que Abrir Dados?** Disponível em: <http://opendatahandbook.org/guide/pt_BR/why-open-data/>. Acesso em: 09 out. 2014.

OPENGOVDATA. **The Annotated 8 Principles of Open Government Data**. Disponível em: <<http://opengovdata.org/>>. Acesso em: 23 out. 2015.

PACHECO, Roberto Carlos dos Santos et al. **Plataforma de Gestão Estratégica à Governança Pública em CT&I**. In: CONGRESSO ABIPTI, 1., 2012, Brasília. Anais... Brasília: S.i., 2012. p. 17 - 33.

PAULO, Rede Nossa São. **Programa Cidades Sustentáveis**. Disponível em: <<http://www.cidadessustentaveis.org.br/institucional>>. Acesso em: 14 out. 2015.

PANIAN, Zeljko. Some Practical Experiences in Data Governance. **World Academy Of Science, Engineering And Technology**, S.i., v. 6, n. 2, p.939-946, out. 2010.

PIZZOL, Leandro dal. **USO DA WEB DE DADOS COMO FONTE DE INFORMAÇÃO NO PROCESSO DE INTELIGÊNCIA COMPETITIVA SETORIAL**. 2014. 138 f. Dissertação (Mestrado) - Curso de Engenharia e Gestão do Conhecimento, Ppegc, Ufsc, Florianópolis, 2014. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2015/04/Leandro-Dal-Pizzol.pdf>>. Acesso em: 22 set. 2015.

PROJECT OPEN DATA. **Common Core Metadata Schema v1.0**. 2015. Disponível em <<https://project-open-data.cio.gov/schema/>>. Acesso em 02 jan 2016.

QUAN, Eilen. **MINNESOTA METADATA GUIDELINES FOR DUBLIN CORE METADATA**. Minnesota Department of Natural Resources. 2000.

RAUTENBERG, Sandro. **MODELO DE CONHECIMENTO PARA MAPEAMENTO DE INSTRUMENTOS DA GESTÃO DO CONHECIMENTO E DE AGENTES COMPUTACIONAIS DA ENGENHARIA DO CONHECIMENTO BASEADO EM ONTOLOGIA**. 2009. 238 f. Tese (Doutorado) - Curso de Engenharia e Gestão do Conhecimento, Ppegc, Ufsc, Florianópolis, 2009. Disponível em: <<http://btd.egc.ufsc.br/wp-content/uploads/2010/06/Sandro-Rauntebrg.pdf>>. Acesso em: 22 set. 2015.

RAUTENBERG, Sandro et al. Ferramenta ontoKEM: uma contribuição à Ciência da informação para desenvolvimento de ontologias. **Perspectivas em Ciência da Informação**, [s.i.], v. 15, n. 1, p.239-258, jan. 2010. Disponível em: <<http://www.scielo.br/pdf/pci/v15n1/14.pdf>>. Acesso em: 15 ago. 2015.

RAUTENBERG, Sandro; TODESCO, José L.; GAUTHIER, Fernando A. O.. Processo de desenvolvimento de ontologias: uma proposta e uma ferramenta. **Revista Tecnológica**, Fortaleza, v. 30, n. 1, p.133-144, jun. 2009.

REINO UNIDO. UNITED KINGDOM (UK) GOVERNMENT. **Opening Up Government**. 2015. Disponível em: <<http://data.gov.uk/about>>. Acesso em: 15 jul. 2015.

RESMINI, Andrea; ROSATI, Luca. A Brief History of Information Architecture. **Journal Of Information Architecture**. [S. I.], p. 33-46. jan. 2012. Disponível em: <<http://journalofia.org/volume3/issue2/03-resmini/>>. Acesso em: 15 dez. 2014.

REINO, Lucas. Informação georreferenciada no Twitter durante as manifestações do 15 de março de 2015.6 ° Simpósio Internacional de CyberJornalismo.2015.

ROVER, Aires José. O governo eletrônico e a inclusão digital: duas faces da mesma moeda chamada democracia. In: ROVER, Aires José(ed). **Inclusão digital e governo eletrônico**. Zaragoza: Prensas Universitárias de Zaragoza, Lefis series 3, 2008, p. 9 - 34.

RIBEIRO, Cláudio José Silva; PEREIRA, Durval Vieira. A publicação de dados governamentais abertos: proposta de revisão da classe sobre Previdência Social do Vocabulário Controlado do Governo Eletrônico. **Transinformação**, [s.l.], v. 27, n. 1, p.73-82, 2015. FapUNIFESP (SciELO). DOI: 10.1590/0103-37862015000100007.

RIVERA, Janessa; MEULEN, Rob van Der. **Gartner Says the Internet of Things Installed Base Will Grow to 26 Billion Units By 2020**. 2013. Disponível em: <<http://www.gartner.com/newsroom/id/2636073>>. Acesso em: 02 jul. 2014.

ROBLES, Rosslin John; KIM, Tai-hoon. Review: Context Aware Tools for Smart Home Development. **International Journal Of Smart Home**. Daegu, p. 1-12. jan. 2010. Disponível em: <http://www.sersc.org/journals/IJSH/vol4_no1_2010/1.pdf>. Acesso em: 23 set. 2014.

SALVADOR, Valéria Farinazzo Martins et al. Qualidade de dados para

gestão de conhecimento na área de saúde. In: CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE, 10., 2006, Florianópolis. **Anais...** . Florianópolis: S.i., 2006. p. 32 - 38. Disponível em:

<https://www.researchgate.net/publication/255631635_Qualidade_de_Dados_para_Gestao_de_Conhecimento_na_Area_de_Saude>. Acesso em: 04 jan. 2016.

SANTOS, Flávio Marcelo Risuenho dos. **Avaliação da qualidade de informação dos controles de eventos de falha e manutenção de equipamentos industriais**. 2009. 117 f. Dissertação (Mestrado) - Curso de Pós Graduação em Engenharia e Gestão do Conhecimento, Universidade Federal de Santa Catarina, Florianópolis, 2009.

SANTOS, Isabel Maria Francisca dos. **Uma proposta de governança de dados baseada em um método de desenvolvimento de arquitetura empresarial**. 2010. 128 f. Dissertação (Mestrado) - Curso de Programa de Pós Graduação em Informática, Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2010.

SINT, Rolf et al. Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis. In: **SEMWiki - 4TH SEMANTIC WIKI WORKSHOP (SEMWiki 2009) AT THE 6TH EUROPEAN SEMANTIC WEB CONFERENCE (ESWC 2009)**, 4., 2009, Hersonissos. Proceedings..Grécia: Eswc, 2009. p. 1 - 20. Disponível em: <<http://ceur-ws.org/Vol-464/paper-14.pdf>>. Acesso em: 15 set. 2015.

SCHMITT, Maurílio Tiago Brüning; DOROW, Patrícia Fernanda; SOUZA, João Artur de. Aplicação da visão sistêmica e do modelo CESM no desenvolvimento da Inteligência Competitiva. In: **CONGRESSO BRASILEIRO DE GESTÃO DO CONHECIMENTO**, 11., 2012, São Paulo. **Anais...** . São Paulo: Km Brasil 2012, 2012. p. 121 - 132. Disponível em: <<http://www.sbgc.org.br/sbgc/kmbrasil-2012/anais/pdf/TC27.pdf>>. Acesso em: 07 jun. 2015.

SCHRIER, Bill. Government Open Data: Benefits, Strategies, and Use. **Spring**, Washington, v. 4, n. 1, p.12-27, fev. 2014.

SHADBOLT, Nigel et al. Linked Open Government Data: Lessons from

Data.gov.uk. **Ieee Computer Society**. Southampton, p. 16-24. jun. 2012.

SOUZA,M.;VENDRUSCULO, L.; MELO, G. **Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core**. Ci. Inf., Brasília, v. 29, n. 1, p. 93-102, jan./abr. 2000.

TAIT, Tania Fatima Calvi; PACHECO, Roberto C. S.; ABREU, Aline França de. Arquitetura de Sistemas de Informação: Evolução e Análise Comparativa de Modelos. **Produção**, [s.i.], v. 9, n. 1, p.55-64, fev. 1999. Disponível em: <<http://www.scielo.br/pdf/prod/v9n1/v9n1a06.pdf>>. Acesso em: 06 jul. 2015.

TEIXEIRA, Thiciane Mary Carvalho; VALENTIM, Marta Lígia Pomim. Estratégias para Disseminação do Conhecimento Organizacional: o papel da arquitetura da informação; Estrategias para la Diseminación del Conocimiento Organizacional. **Informação & Informação**, [s.l.], v. 17, n. 3, p.165-180, 11 jan. 2013. Universidade Estadual de Londrina. DOI: 10.5433/1981-8920.2012v17n3p165. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/12110/pdf>>. Acesso em: 02 ago. 2015.

THE OPEN GROUP. **The Open Group**: Leading the development of open, vendor-neutral IT standards and certifications. Disponível em: <<http://www.opengroup.org/aboutus>>. Acesso em: 08 jul. 2015.

TOWNSEND,A.**Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia**. W. W. Norton & Company. 2013.

TWITTER, Engineering. **200 million Tweets per day**. 2011. Disponível em: <<https://blog.twitter.com/2011/200-million-tweets-day>>. Acesso em: 18 ago. 14.

TWITTER DEVELOPERS. The Search API. 2015. Disponível em <<https://dev.twitter.com/rest/public/search>>. Acesso em: 31. jan 2015..

US GOVERNMENT. **U.S. Government's Open Data**. 2015. Disponível em: <<http://www.data.gov/>>. Acesso em: 15 jul. 2015.

VILAJOSANA, Ignasi et al. Bootstrapping smart cities through a self-sustainable model based on big data flows. **IEEE Communications**

Magazine, Estados Unidos, v. 51, n. 6, p.128-134, jun. 2013. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&tp;=&arnumber=6525605&queryText;=vilajosana+big+data+bootstrapping>>. Acesso em: 12 jul. 2014.

VLACHEAS, Panagiotis et al. Enabling smart cities through a cognitive management framework for the internet of things. **Ieee Communications Magazine**, New York, v. 51, n. 6, p.102-111, jun. 2013. Institute of Electrical & Electronics Engineers (IEEE). DOI: 10.1109/mcom.2013.6525602.

WALKER, Michael. Data Veracity. **Data Science Central**. 2012. Disponível em : <<http://www.datasciencecentral.com/profiles/blogs/data-veracity>>. Acesso em 22 set. 2015.

WALT, van Der; DUTOIT, A.s.a.. Developing an information model for an enterprise: a South African case study. **South African Journal**. South Africa, p. 1-1. fev. 2006. Disponível em: <<http://www.sajim.co.za/index.php/SAJIM/article/viewFile/663/781>>. Acesso em: 17 ago. 2015.

WANG, Shouhong. Modeling information architecture for the organization. **Information & Management**. Canadá, p. 303-315. maio 1997.

WAZE MOBILE. **WAZE. DERROTANDO O TRÂNSITO, JUNTOS**. Disponível em: <<https://www.waze.com/pt-BR>>. Acesso em: 14 set. 2015.

WEB FOUNDATION. **Web Index data 2012**. 2012. Disponível em: <<http://legacy.thewebindex.org/wp-content/uploads/2012/09/05-webindex-dataset.html>>. Acesso em: 08 jul. 2015.

WHITE, Martin. Information architecture. **Emerald Group Publishing Limited**. [S. I.], p. 218-219. jan. 2004. Disponível em: <<http://www.emeraldinsight.com/toc/el/22/3>>. Acesso em: 15 dez. 2014.

WOOLLASTON, Victoria. **Revealed, what happens in just ONE minute on the internet: 216,000 photos posted, 278,000 Tweets and 1.8m Facebook likes**. 2013. Disponível em: <<http://www.dailymail.co.uk/sciencetech/article-2381188/Revealed->

happens-just-ONE-minute-internet-216-000-photos-posted-278-000-Tweets-1-8m-Facebook-likes.html>. Acesso em: 12 ago. 2014.

W3C.**Introduction to SKOS. 2015.** Disponível em <<https://www.w3.org/2004/02/skos/intro>>. Acesso em: 03 jan 2016.

ZACHMAN, John A.. **The Zachman Framework For Enterprise Architecture:** Primer for Enterprise Engineering and Manufacturing. [s.i]: Zachman International, 2003. 15 p.

ZYGIARIS, Sotiris. Smart City Reference Model: Assisting Planners to Conceptualize the Building of Smart City Innovation Ecosystems. **Journal Of The Knowledge Economy**, [s. l.], v. 4, n. 2, p.217-231, 8 mar. 2012. Springer Science + Business Media. DOI: 10.1007/s13132-012-0089-4.